

**APPLICATION**

**FOR**

**UNITED STATES LETTERS PATENT**

**TITLE: INTERACTION-ACTIVATED PROTEINS**

**INVENTORS: ROBERT F. BALINT  
JENG-HORNG HER**

009760-50732560

**CERTIFICATE OF EXPRESS MAILING**

"Express Mail" Label No: EK113060819US

Date of Deposit: 15 March, 2000

I hereby certify under 37 C.F.R. 1.10 that this correspondence is being deposited with the United States Postal Service as "Express Mail Post Office to Addressee" with sufficient postage on the date indicated above and is addressed to the Assistant Commissioner of Patents, Washington, D.C. 20221.

Jennifer Wahlsten  
(Signature)

Jennifer Wahlsten  
(Printed Name)

## **INTERACTION-ACTIVATED PROTEINS**

### **CROSS REFERENCE TO RELATED APPLICATIONS**

This application claims the benefit of US Provisional Application No. 60/124,339, filed March, 15, 1999, and US Provisional Application No. 60/135,926, filed May 25, 1999, and US Provisional Application No. 60/175,968, filed January 13, 2000, which disclosures are hereby incorporated by reference.

### **GOVERNMENT LICENSE RIGHTS**

The U. S. government has a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of grant No. \_\_\_\_\_ for or awarded by \_\_\_\_\_.

## **INTRODUCTION**

### **Technical Field**

The present invention is concerned with detecting interactions between proteins by expressing them as part of a fusion sequence that also encodes for one fragment of a fragment pair that reassembles into a directly detectable protein. The interaction-dependent enzyme association (IdEA) systems of the present invention are exemplified by the bacterial  $\beta$ -lactamases, a large group of structurally-related enzymes which segregate into several groups on the basis of structural homologies and substrate specificities.

### **Background**

Most physiological processes depend on complex networks of cells interacting with one another and their environments, primarily through specific recognition between proteins – from the ligand-mediated assembly of multi-protein complexes at the cell surface, through the labyrinth of intracellular signal transduction cascades, to the assembly of transcription-modulating complexes on the promoters of specific genes. Thus, for most pathological conditions, protein-protein interactions are instrumental and provide a wealth of targets for diagnostic and therapeutic intervention. As a result, new and improved

methods are in constant demand for (1) identifying natural ligands of key participants to study their roles in disease, and (2) developing surrogate ligands for therapeutic intervention and diagnosis. A number of methods have been developed over the years to address each of these goals. The most widely used current methods for identifying natural proteins which interact with a protein-of-interest generally involve screening libraries of expressed cDNAs. A few genes for ligands of proteins-of-interest have been isolated by direct screening of cDNA expression libraries on filters for binding to labeled versions of the protein-of-interest, as in antibody probing (Blackwood and Eisenman, *Science* (1991) 251:1211; Defeo-Jones *et al.*, *Nature* (1991) 352:251). However, a great many important protein interactions are not robust enough for the harshness of such methods, where conditions of interaction are usually far from native. Also, the false positive frequencies of these methods is high, due to the presence of denatured protein in cells which have been fixed to make the target proteins accessible to probes.

A major advance in cDNA screening methodology came with the development of systems in which screenable or selectable cellular phenotypes could be engineered to depend on desired protein interactions within living cells (Fields and Song *Nature* (1989) 340:245; Chien *et al.*, *Proc Natl Acad Sci* (1991) 88:9578; Zervos *et al.*, *Cell* (1993) 72:223; Vojtek *et al.*, *Cell* (1993) 74:205; and Luban *et al.*, *Cell* (1993) 73:1067). The most widely used of these is the yeast "two hybrid" system of Fields and Song (1989, *supra*). This system takes advantage of the "modularity" of many functional domains in proteins which allows the linking of functions to be manipulated. This is particularly true for transcriptional activators, in which an activation domain which interacts with the core transcription complex is "homed" to specific genes by a sequence-specific DNA-binding domain. For many transcriptional activators these domains can function independently, and in fact are often in separate, interacting subunits. In the yeast two-hybrid system, the "bait" protein is expressed as a fusion with a *cis*-element sequence-specific DNA-binding domain, and cDNAs are expressed as fusions with a transactivation domain. When, and only when, these two domains are brought together by interaction of a cDNA product with the "bait" protein, can the reporter gene be expressed, since its transcription is dependent on transactivation from the *cis*-element. Reporters can be either screenable (e.g.,  $\beta$ -galactosidase for color) or selectable (e.g., *HIS3* for growth in the absence of histidine).

Variations of this system have been successfully employed to identify a number of

important protein-protein interactions (Chien *et al.*, 1991, *supra*; Zervos *et al.*, 1993, *supra*; Vojtek *et al.*, 1993, *supra*; and Luban *et al.*, 1993, *supra*; Bartel *et al.*, *Nature Genetics* (1996) 2:72; Fromont-Racine *et al.*, *Nature Genetics* (1997) 3:277; Xu *et al.*, *Proc Natl Acad Sci* (1997) 94:12473). In spite of its success, however, the original yeast two-hybrid system has serious drawbacks for the high-throughput applications required to accelerate pharmaceutical target discovery from genomics. The fundamental limitation with this system is that many steps are required between the test interaction and the generation of a selectable phenotype. Each such step presents an opportunity for non-specific interaction to raise the false positive background, and for dissociation to allow bona fide interactors to be missed. The false positive problem is exacerbated by the highly combinatorial nature of the transcription machinery and the abundance of protein domains encoded in cDNA libraries which can interact with one or more components of the transcription initiation complex, including transactivator-bound promoter DNA (Bartel *et al.*, *BioTechniques* (1993) 14:920). Another limitation of the original two-hybrid system is that it generally cannot accommodate secreted or membrane proteins and cytoplasmic proteins must be stable in the yeast nucleus.

Recently the two-hybrid concept has been expanded to include other types of protein functionalities for use as protein-protein interaction reporting systems. For example, in the Selective Infective Phage (SIP) system a protein which confers infectivity on filamentous bacteriophage has been fragmented in such a way that it is functional only when the fragments are fused to heterologous interactors (Krebber *et al.*, *J Mol Biol* (1997) 268:607). The interaction is then monitored by its ability to allow phage encoding the interactors to transfer a selectable phenotype to susceptible cells by infection. However, this method also suffers from requiring many low-efficiency steps between the target interaction and the expression of the selectable phenotype by the recipient cell. Also like the two-hybrid system, the efficiency of this system suffers from the fact that most natural protein-protein interactions have affinities in the micromolar range, with half-lives on the order of seconds. When the time delay between interaction and signal generation exceeds this half-life, which it does in these systems, the efficiency of interaction detection declines sharply.

More recently still, the two-hybrid concept has been adapted to proteins which can confer selectable phenotypes directly from protein-protein interactions, with few or no

intervening steps between the target interaction and signal generation. For example, interactors can be fused to variants of the Green Fluorescent Protein of *Aequorea victoria* (GFP), which are capable of detectable fluorescence resonance energy transfer (FRET) when brought into close proximity by the interactors (Cubitt *et al.*, *Trends Biochem* (1995) 20:448). Some enzymes which confer selectable or screenable phenotypes on cells can also be adapted for two-hybrid type protein-protein interaction detection (Rossi *et al.*, *Proc Natl Acad Sci* (1997) 94:8405; Pelletier *et al.*, *Proc Natl Acad Sci* (1998) 95:12141). In this variation, protein interactors are fused to enzyme fragments, which by themselves are inactive. However, when the enzyme fragments are brought together by the interaction of the protein domains to which they are fused, the fragments are able to associate to reconstitute the selectable activity of the enzyme. This is an example of interaction-dependent enzyme activation (IdEA), and it is illustrated in Figure 1. Both IdEA and GFP FRET systems present advantages over previous versions of the two-hybrid concept. For instance, the selectable signal is produced directly from the desired interaction, without any intervening steps which are the main sources of inefficiency in the earlier systems. Such improvements in efficiency and background should make these methods more amenable to high-throughput applications. However, although both IdEA and GFP FRET systems in theory can be set up in both prokaryotic and eukaryotic cells, and either in the cytoplasm or in a secretory pathway to allow interactions to be monitored in natural milieus, they have not. All IdEA systems reported to date have only utilized cytoplasmic enzymes and have only been shown to be operative in that compartment (Rossi *et al.*, 1997, *supra*; Pelletier *et al.*, 1998, *supra*; Karimova *et al.*, *Proc Natl Acad Sci* (1998) 95:5752). Indeed, because of their design, these reported systems would not be expected to function in the secretory pathway or in the bacterial periplasm. Thus, they are not considered useful for monitoring the interactions of secreted proteins.

The most widely used current systems for the detection of extra-cellular protein-protein interactions, namely viral or cellular display systems, are essentially *in vitro* methods with high stringencies of selection and/or high backgrounds. Thus, they are not well suited for high-throughput applications. These systems also usually require the use of a purified known heterologous interactor domain or "bait protein", and are therefore not suitable for multiplex applications where neither heterologous interactor domain of a protein binding pair is known *a priori*, i.e., the combinatorial interaction of two protein libraries with one another for

simultaneous identification of all protein binding pair interactions. One system which does not require bait purification for identification of extra-cellular interactions is the *E. coli* Dimer Detection System (EDDS; Small Molecule Therapeutics, Inc., Monmouth Junction, NJ). Bait proteins for this system are restricted to type I membrane receptors which have single transmembrane domains and require simple dimerization for signaling. The ecto-domain of the bait receptor is fused to the transmembrane domain and endo-domain of an *E. coli* receptor. When this fusion protein is co-expressed with an expression library in the bacterial periplasm, ligands for the receptor can be identified by their ability to dimerize the receptor and induce expression of a selectable phenotype. However, this system suffers from the same limitation as the yeast two-hybrid and SIP systems, namely, that multiple steps between interaction and phenotype cause severe loss of efficiency due to high false positive and false negative rates.

It is therefore of interest to develop IdEA systems capable of simultaneous detection of multiple interactions between extra-cellular as well as intracellular proteins in a high throughput format.

#### Relevant Literature

USPN 5,585,245 discloses a ubiquitin-based protein sensor complementation system where binding of two predetermined proteins of a binding pair is detected as specific proteolysis of ubiquitin by ubiquitinases. PCT publication WO 98/44350 discloses a reporter subunit complementation system employing fusion proteins of  $\beta$ -galactosidase subunits. PCT publication WO 98/34120 discloses a protein fragment complementation system employing dihydrofolate reductase.

#### SUMMARY

Compositions and methods are provided for identifying interactions between polypeptides using an interaction-dependent protein association system. The system is characterized by using fragment pairs comprised of a first and a second member that functionally reassemble into a marker protein having a directly detectable signal, such as a visible phenotypic change or antibiotic resistance. The fragment complementation system of the present invention involves co-expression in a host cell of a first and a second oligopeptide, where each is a fusion protein separated by a flexible polypeptide linker with

a member of the marker protein fragment pair. Binding of the first oligopeptide to the second oligopeptide results in the functional reconstitution of the fragment pair into a marker protein, and the interacting first and second oligopeptides are identified by isolating and sequencing plasmids from a host cell that displays a directly detectable signal indicative of the marker protein. Functional reconstitution of the fragment pairs into a marker protein can be enhanced by including elements such as a cysteine residue or a randomly encoded peptide of from 3-12 amino acids at or near the break-point termini of the fragment pair member, or by introducing 1-3 codon changes within the nucleotide sequence encoding for a member of a fragment pair. The invention also provides for efficient methods of finding functional fragment pairs of a marker protein that involve identifying functional break-points within flexible loops using tertiary or secondary structural information. The interaction-dependent protein activation systems of the present invention find particular use in identifying immunoglobulin epitopes, polypeptide sequences that bind to extracellular proteins, and inhibitors of phosphorylation-regulated signal transducer proteins. The systems also find use in allowing single antibiotic selection of cells transformed to express genes for multiple traits and for targeted and localized activation of derivitized anti-tumor prodrugs.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

**Figure 1.** Mechanism for Interaction-dependent Enzyme Activation (IdEA). Interaction-dependent fragment complementation requires enzyme  $\alpha$  and  $\omega$  fragments which can refold to form active enzyme when and only when they are brought together by an interaction of heterologous domains fused to their termini.

**Figure 2.** Nucleotide coding sequence for the mature form of TEM-1  $\beta$ -lactamase and the encoded amino acid sequence (Sutcliffe, *Proc Natl Acad Sci* (1978) 75:3737). From the sequence for plasmid pBR322 (SYNPBR322), Genbank accession no. J01749. The break-points between the  $\alpha$  and  $\omega$  fragments at residues Asn52/Ser53, Glu63/Glu64, Gln99/Asn100, Pro174/Asn175, Glu197/Leu198, Lys215/Val216, Ala227/Gly228 and Gly253/Lys254 are indicated.

**Figure 3.** Three-dimensional structure of mature TEM-1  $\beta$ -lactamase. Rendering of the x-ray crystal structure of Jelsch *et al.* (*Proteins Struct Funct* (1993) 16:364ff), using red and blue solid ribbons to show  $\alpha$ -helix and  $\beta$ -sheet, respectively. The molecule is oriented to emphasize the two-domain structure ( $\alpha$ - $\omega$  and  $\mu$ ). The active site nucleophile, Ser70, is shown as a ball-and-stick model.

**Figure 4.** Three-dimensional representation of interaction-dependent activation of  $\beta$ -lactamase by fragment complementation. Docking of TEM-1  $\alpha$ 197 and  $\omega$ 198 fragments by the interaction of the hetero-dimerizing helices from the fos and jun subunits of the AP-1 transcription activator allows re-folding of the fragments into the active conformation of the enzyme (compare with Figure 3).

**Figure 5.** Structures of some anti-cancer drugs and their cephalosporin prodrugs. YW-200 and YW-285 are a DNA-binding tri-indole and its cephalosporin prodrug (Wang *et al.*, 1998, US Patent 5,843,937)

**Figure 6.** Vectors and strategy for the expression of heterologous proteins as fusions to the  $\alpha$ 197 and  $\omega$ 198 fragments of TEM-1  $\beta$ -lactamase for interaction-dependent  $\beta$ -lactamase activation by fragment complementation. Vector pAO1 is a high-copy pUC119-based phagemid for expression of  $\omega$ 198 fusions and free ligands from dicistronic transcripts, which can be rescued as phage for quantitative introduction into host cells by high-multiplicity infection. Vector pAE1 is a low-copy p15A replicon with a strong promoter for expression of  $\alpha$ 197 fusions at comparable or higher levels than expression from the pAO1 vector. Tripeps are 12-mer peptides inserted into the active site of thioredoxin. Tripep-trx libraries are random tri-peptides at the N-terminus of thioredoxin with an intervening Gly<sub>4</sub>Ser linker. ScFv, single-chain antibody Fv fragment. LC-CH1, antibody fragment composed of light chain and first constant region of heavy chain. VL, antibody light chain variable region. *lac* prom, lactose operon promoter. SP, signal peptide. (Gly<sub>4</sub>Ser)<sub>3</sub>, flexible 15-mer linker. pUC ori, p15A ori, plasmid origins of replication. f1 ori, filamentous phage origin of replication. *cat*, chloramphenicol resistance gene. m.o.i., multiplicity of infection. *trc* prom, fusion promoter from tryptophan and lactose operons. *tt*, transcription terminator. *kan*, kanamycin resistance gene. Vector sizes in base pairs



Sub 2004  
(bp) do not include interactors.

**Figure 7.** TEM-1  $\beta$ -lactamase fragment complementation by interaction between representative single-chain antibody Fv fragment (scFv) and thioredoxin-scaffolded peptide (Trx). The N-terminal  $\beta$ -lactamase fragment,  $\alpha$ 197 ( $\alpha$ ), is colored red. The C-terminal fragment,  $\omega$ 198 ( $\omega$ ), is colored blue. TEM-1, thioredoxin, and the scFv were rendered from published structures. The peptide and the linkers were drawn in.

**Figure 8.** TEM-1  $\beta$ -lactamase fragment complementation by interaction between the CD40 extra-cellular domain (CD40) and a thioredoxin-scaffolded peptide (Trx). The N-terminal  $\beta$ -lactamase fragment,  $\alpha$ 197 ( $\alpha$ ), is colored red. The C-terminal fragment,  $\omega$ 198 ( $\omega$ ), is colored blue. TEM-1, thioredoxin, and the scFv were rendered from published structures. The peptide and the linkers were drawn in.

**Figure 9.** Vectors and protocol for construction of a multiplex protein-protein interaction library using interaction-dependent  $\beta$ -lactamase fragment complementation systems. Expressed sequence (ES), i.e., random-primed cDNA libraries, are subcloned into phagemid vectors for expression as fusions to the  $\beta$ -lactamase  $\alpha$  and  $\omega$  fragments, via the flexible linker (Gly<sub>4</sub>Ser)<sub>3</sub>. The vectors encode a peptide epitope tag, such as the 12-residue Myc tag, at the C-terminus of the ES. When co-expressed with anti-Tag scFv, such as anti-myc 9E10, fused to the other fragment, the ES libraries can be selected for  $\beta$ -lactamase activity driven by the Tag-anti-Tag interaction, which will require stable expression of the ES fragment. The resultant libraries, enriched for stable expressors of autonomously folding domains (AFD), may then be rescued as phage and co-infected into male cells for selection of interacting AFD pairs (Multiplex Interaction Library). The AFD libraries can also be co-infected with scFv libraries, antibody light chain variable region libraries (VL), or peptide libraries displayed on thioredoxin (trx-peptide) for simultaneous selection of binding proteins for each AFD (Multiplex Antibody/Peptide Binder Selection). See legends to Figures 6 and 10 for identification of other abbreviations.

**Figure 10.** Abbreviated output of the PredictProtein Program for prediction of secondary structure and solvent exposure for NPTII (Rost and Sander, 1993, 1994). The top line

shows the amino acid sequence in single letter code. The second and third lines show secondary structure prediction. H, helix; E, strand; L, loop. The fourth line shows a measure of reliability on a scale from 1 to 10, with 10 being highest. The fifth line shows solvent accessibility – e, exposed; b, buried. The bottom line shows a measure of reliability for solvent accessibility on a scale of 1 to 10, with 10 being highest. Ten regions of the sequence predicted to have little secondary structure and to be exposed to solvent are indicated by underlining as potential sites for productive fragmentation.

**Figure 11.** Expression vectors for production of  $\beta$ -lac $\alpha$ 253 and  $\beta$ -lac $\omega$ 254 fusion proteins with scFv. Arrows denote translation start sites. T7 prom, bacteriophage T7 promoter; SP, pelB signal peptide; scFv is comprised of VH (antibody heavy chain variable region), (Gly<sub>4</sub>Ser)<sub>3</sub> (15-mer flexible linker), and VL (antibody light chain variable region); *kan*, kanamycin resistance; His<sub>6</sub>, hexa-histidine tag for metal ion affinity purification; *lacI<sup>q</sup>*, high-affinity *lac* operon repressor mutant; f1 ori, phage origin of replication.

#### **BRIEF DESCRIPTION OF THE SPECIFIC EMBODIMENTS**

Methods and compositions are provided for an interaction-dependent protein activation system useful in detecting an interaction between a first protein and a second target protein. The method detects the interaction of a first known or unknown interactor domain with a second unknown interactor domain by bringing into close proximity members of a fragment pair of a marker protein, such that the parent marker protein is reassembled to its original functionality, and such that reassembly requires the prior interaction of the heterologous interactor domains. The system is characterized by N-terminal and C-terminal fragment members that comprise fragment pairs which are derived from, and can functionally reassemble into a marker protein that provides for a directly detectable signal that does not involve downstream steps necessary for recognition. For example, a marker protein of interest for the instant invention functions of itself to produce a selectable signal such as a visible phenotypic change or antibiotic resistance.

The fragment pairs are used in methods that involve the co-expression of a first and a second oligopeptide sequence, in which the first oligopeptide sequence is a fusion protein comprised of in the direction of translation, an N-terminal fragment fused through a break-point terminus to a flexible polypeptide linker and a first interactor domain, and the second

oligopeptide sequence is a fusion protein comprised of in the direction of translation, a second interactor domain and a flexible polypeptide linker fused through a break-point terminus to a C-terminal fragment. The flexible polypeptide linker separates the fragment domain from the interactor domain and allows for their independent folding. The linker is  
5 optimally 15 amino acids or 60 Å in length ( $\sim 4$  Å per residue) but may be as long as 30 amino acids but preferably not more than 20 amino acids in length. It may be as short as 3 amino acids in length, but more preferably is at least 6 amino acids in length. To ensure flexibility and to avoid introducing steric hindrance that may interfere with the independent folding of the fragment domain and the interactor domain, the linker should be comprised  
10 of small, preferably neutral residues such as Gly, Ala and Val, but also may include polar residues that have heteroatoms such as Ser and Met, and may also contain charged residues.

The first interactor domain is a known or unknown protein or protein fragment that binds directly or indirectly to a second target interactor domain that is an unknown protein or protein fragment and either or both the first and second interactor domain can be a member of a library. The interactor domain libraries are preferably constructed from cDNA, but may also be constructed from, for example, synthetic DNA, RNA and genomic DNA. When combining the first and second oligopeptide sequences, the reconstitution of the N-terminal and C-terminal fragments into the marker protein requires the prior  
15 interaction of the first and second interactor domains. Bound interactor domains are identified by expressing a functionally reconstituted marker protein, and then the nucleotide sequences encoding for bound interactor domains or the bound interactor domains themselves are characterized by methods including electrophoresis, polymerase chain reaction (PCR), nucleotide and amino acid sequencing and the like.  
20

Advantages of the present invention over previously disclosed fragment complementation systems include a reporter protein that provides for a directly detectable signal upon reassembly, and background levels of 1 in  $10^6$  or less. Additionally, the invention provides for rationally incorporated enhancement modifications to the fusion oligopeptides that increase the functional activity of the reconstituted protein to wild-type  
25 levels by improving folding and reassembly of the fragments into the parent protein, while at the same time maintaining dependence on the interactor domains for reassembly.

The interaction-dependent enzyme activation system of the subject invention may be

used to detect *in vitro* protein interactions, such as in cell lysates, or the interactions of intracellular or extracellular proteins of a host cell. For evaluating interactions between extracellular proteins, the first and second fusion oligopeptides can be expressed with a signal peptide. In bacterial host cells, for example, an N-terminal signal peptide can provide for translocation of the fusion oligopeptides to the periplasm. The combined lengths of the N-terminal fragment and the C-terminal fragment may be discontinuous with residues around the break-point deleted, contiguous, or overlapping with residues around the break-point repeated, thereby comprising from 90% to 110% of the total length of the parent protein. Break-point termini are herein defined as the C-terminus of the N-terminal fragment and the N-terminus of the C-terminal fragment.

The subject invention provides for enhancing the performance of the reassembled parent protein by introducing at least one of the following modifications, including: i) a randomly-encoded peptide of 3-12 amino acids between the break-point terminus of each fragment and the flexible polypeptide linker; ii) a randomly-encoded peptide of 3-12 amino acids expressed separately as a fusion to the N-terminus of a thioredoxin with an intervening flexible linker; iii) a cysteine residue encoded at or within 5 amino acid positions of the break-point and between the break-point terminus of each fragment and the flexible polypeptide linker so that a disulfide bond can form between the members of a fragment pair, and iv) 1-3 codon changes within a member of a fragment pair introduced, for example, by PCR amplification of a nucleotide sequence encoding for a member of a fragment pair under error-prone conditions, to enhance the folding stability of a functionally reconstituted marker protein.

The invention is also directed to plasmids containing expression cassettes constructed to express fusion oligopeptides comprised of a fragment domain and an interactor domain. The expression cassettes for the N-terminal and C-terminal fragment pair members are designed with their components in different sequential orders. For the C-terminal fragment pair member, the expression cassette will comprise as operably linked components in the direction of transcription nucleotide sequences encoding for (i) a promoter functional in a host cell, (ii) a polypeptide interactor domain, (iii) a flexible polypeptide linker and (iv) a C-terminal fragment of a marker protein that provides for a directly selectable phenotype. The expression cassette for the N-terminal fragment pair member will comprise as operably linked components in the direction of transcription

nucleotide sequences encoding for (i) a promoter functional in a host cell, (ii) an N-terminal fragment of a marker protein that provides for a directly selectable phenotype, (iii) a flexible polypeptide linker and (iv) a polypeptide interactor domain. The invention is also concerned with host cells that contain plasmids having the sequences of the above-described expression cassettes.

Appropriate host cells for application of the subject invention include both eukaryotic cells, such as mammalian, yeast and plant cells, and prokaryotic cells, such as bacterial cells. A variety of prokaryotic expression systems can be used to express the fusion oligopeptides of the subject invention. Expression vectors can be constructed which contain a promoter to direct transcription, a ribosome binding site, and a transcriptional terminator. Examples of regulatory regions suitable for this purpose in *E. coli* are the promoter and operator region of the *E. coli* tryptophan biosynthetic pathway as described by Yanofsky (1984) *J. Bacteriol.*, 158:1018-1024 and the leftward promoter of phage lambda ( $P_L$ ) as described by Herskowitz and Hagen, (1980) *Ann. Rev. Genet.*, 14:399-445. Vectors used for expressing foreign genes in bacterial hosts generally will contain a sequence for a promoter which functions in the host cell. Plasmids useful for transforming bacteria include pBR322 (Bolivar, *et al.*, (1977) *Gene* 2:95-113), the pUC plasmids (Messing, (1983) *Meth. Enzymol.* 101:20-77, Vieira and Messing, (1982) *Gene* 19:259-268), pCQV2 (Queen, *ibid.*), and derivatives thereof. Plasmids may contain both viral and bacterial elements. Methods for the recovery of the proteins in biologically active form are discussed in U.S. Patent Nos. 4,966,963 and 4,999,422, which are incorporated herein by reference. See Sambrook, *et al.* (In *Molecular Cloning: A Laboratory Manual*, 2<sup>nd</sup> Ed., 1989, Cold Spring Harbor Laboratory Press, Cold Spring Harbor) for a description of other prokaryotic expression systems.

For expression in eukaryotes, host cells for use in practicing the present invention include mammalian, avian, plant, insect, and fungal cells. As an example, for plants, the choice of a promoter will depend in part upon whether constitutive or inducible expression is desired and whether it is desirable to produce the fusion oligopeptides at a particular stage of plant development and/or in a particular tissue. Expression can be targeted to a particular location within a host plant such as seed, leaves, fruits, flowers, and roots, by using specific regulatory sequences, such as those described in USPN 5,463,174, USPN 4,943,674, USPN 5,106,739, USPN 5,175,095, USPN 5,420,034, USPN 5,188,958, and

USPN 5,589,379.

Where the host cell is a yeast cell, transcription and translational regions functional in yeast cells are provided, particularly from the host species. The transcriptional initiation regulatory regions can be obtained, for example from genes in the glycolytic pathway, such as alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase (GPD), phosphoglucosomerase, phosphoglycerate kinase, etc. or regulatable genes such as acid phosphatase, lactase, metallothionein, glucoamylase, etc. Any one of a number of regulatory sequences can be used in a particular situation, depending upon whether constitutive or induced transcription is desired, the particular efficiency of the promoter in conjunction with the open-reading frame of interest, the ability to join a strong promoter with a control region from a different promoter which allows for inducible transcription, ease of construction, and the like. Of particular interest are promoters which are activated in the presence of galactose. Galactose-inducible promoters (GAL1, GAL7, and GAL10) have been extensively utilized for high level and regulated expression of protein in yeast. (Lue *et al.*, (1987) *Mol. Cell Biol.* 7:3446; Johnston, (1987) *Microbiol. Rev.* 51:458).

The invention also provides for efficient methods of identifying functional fragment pairs of a marker protein of interest that involves preparing a multiplicity of fragment pair members with break-point termini within a solvent exposed loop or a flexible loop defined by tertiary or secondary structure analysis to obtain a fragment pair library. The fragment pair members are expressed in a multiplicity of host cells, and the host cells exhibiting the directly detectable signal associated with the marker protein of interest are isolated as indicative of containing fragment pair members that functionally reconstitute the marker protein. Plasmids containing expression cassettes coding for the fragment pair members are then sequenced to identify functional fragment pairs. To aid in the identification of functional fragment pair members of a marker protein of interest, the fragment pair members can be expressed as fusion proteins with interactor domains known to bind to each other, such as the fos and jun transcription factors that associate through a leucine zipper interaction. The sequences encoding the hetero-dimerizing helices of the fos and jun transcription factors are sufficient to use as effective interactor domain for this purpose.

The systems and methods of the subject invention find particular use in identifying epitopes recognized by immunoglobulin molecules, polypeptide sequences that bind to extracellular domains of a transmembrane protein, inhibitors of phosphorylation-regulated

signal transducer proteins, and interaction between oligopeptides of two different proteomes. For the identification of epitopes, first and second fusion oligopeptides comprised of a fragment domain and an interactor domain are expressed in a host cell where the first fusion oligopeptide has an interactor domain comprised of a randomly encoded peptide inserted into the active site of a thioredoxin protein and the interactor domain of the second fusion oligopeptide is comprised of a single-chain variable region (scFv) or antibody light chain variable region (VL). A similar strategy is followed for identifying polypeptide sequences that interact with the extracellular domain of a transmembrane protein, where the first interactor domain is comprised of a randomly encoded peptide inserted into the active site of a thioredoxin protein and the second interactor domain is comprised of a transmembrane protein. Identification of inhibitors of a phosphorylation-regulated signal transduction protein involves expressing a first fusion oligopeptide with a first interactor domain comprised of a phosphorylation-regulated signal transduction protein, such as Her-2/neu, and a second fusion oligopeptide with a second interactor domain comprised of a scFv or antibody light chain variable region that only binds to the unphosphorylated signal transduction protein. Inhibitory compounds are identified from host cells that change color in the presence of a chromogenic  $\beta$ -lactamase substrate. For identifying or monitoring polypeptide-polypeptide interactions between the members of two different proteomes, members of a first and second cellular expression library comprise the first and second interactor domain, respectively, of a fusion oligopeptide. The expression library is preferably a cDNA library, but may also be constructed from synthetic nucleotides to screen randomly generated polypeptides. A library of particular application for the present invention should represent all the protein members of a proteome of interest. Libraries derived from nucleotide sequences that all members of a total protein population (i.e. a proteome) of interest may be isolated from a host cell such as a prokaryotic or a eukaryotic cell, or also from a viral host. Viral hosts that encode for oncogenes are of particular interest. Mammalian tumor cells, immune cells and endothelial cells also provide proteomes of particular interest for the subject invention.

The invention also finds use in selecting with a single marker protein the incorporation of multiple genetic traits in a host cell, where detectable expression of a functionally reassembled marker protein is indicative of co-expression of multiple genes that encode for individual traits in a host. Finally, the invention provides therapeutic utility

in a method for specifically activating derivitized prodrugs in the vicinity of a target organ in a host, where each member of a marker protein fragment pair is expressed as a fusion protein with individual immunoglobulin molecules that recognize neighboring but non-overlapping epitopes on a target protein. Binding of both antibodies to the target protein allows functional reconstitution of the marker protein which then activates subsequently administered prodrug only in the vicinity of a target organ.

The invention is exemplified by the antibiotic resistance enzyme, TEM-1  $\beta$ -lactamase, although fragment pairs of other enzymes that provide for antibiotic resistance are included in the present invention, including: aminoglycoside phosphotransferases, particularly neomycin phosphotransferase, chloramphenicol acetyl transferase, and the tetracycline resistance protein described by Backman and Boyer (*Gene* (1983) 26:197). Other proteins that can directly elicit a visible phenotypic change such as a color change or fluorescence emission also are applicable to the subject invention. Examples of such proteins include  $\beta$ -galactosidase and green fluorescent protein (GFP) or other related fluorescent proteins.

The TEM-1  $\beta$ -lactamase of *E. coli* is the 264 amino acid product of the ampicillin resistance gene of plasmid pBR322 (Sutcliffe, 1978, *supra*), the nucleotide sequence of which is shown in Figure 2 along with the encoded amino acid sequence. TEM-1 is the archetype member of the homologous Class A  $\beta$ -lactamases, or penicillinases. Its three-dimensional structure is shown in Figure 3 (Jelsch *et al.*, *Proteins Struct Funct* (1993) 16:364ff). The Class A  $\beta$ -lactamases are comprised of two domains. One domain,  $\alpha$ - $\omega$ , is made up of N-terminal and C-terminal sequences, which form an anti-parallel two-helix bundle packed against a flat 5-stranded  $\beta$ -sheet. The inner face of the sheet packs against the other domain ( $\mu$ ), a seven helix bundle with two extended loops and two small  $\beta$ -structures. An outside strand of the  $\beta$ -sheet borders the substrate binding pocket, opposite the catalytic nucleophile, Ser70, and contributes substrate-binding residues. The remainder of the active site residues, including Ser70, are contributed by the  $\mu$  domain. The two domains are connected by two loops: R61-R65 and D214-W229.

The subject invention also provides a method of identifying optimal break-points in a parent protein that provides for a directly detectable signal. A search of the "fragment space" of TEM-1  $\beta$ -lactamase was conducted to identify fragment pairs which could complement for activity only when the break-point termini of the fragments were



genetically fused to hetero-dimerizing helices from the *c-fos* and *c-jun* subunits of the AP-1 transcription factor (Karin *et al.*, *Curr Opin Cell Biol* (1997) 9:240. To do this, libraries of all possible N- and C- terminal fragments of the enzyme were generated by progressive exonucleolytic digestion of the full coding sequence from both termini. Fragments of less than 25 amino acids were considered non-viable. When libraries were constructed with compatible vectors, the fragment sequences co-expressed in the same *E. coli* cells so that each cell expressed a single pair of N- and C- terminal fragments and every possible pair may be represented. For example, for a 100 kDa enzyme there are only  $10^6$  possible N- and C-terminal fragment pairs, so an exhaustive search of the fragment space of most enzymes could be conducted with libraries of a manageable size. An exposed loop was identified by this method between two  $\alpha$ -helices of *E. coli* TEM-1  $\beta$ -lactamase (approximately Thr195 to Ala202, between helices 7 and 8) within which the chain could be broken to produce fragments which could only complement for activity when fused to the *fos* and *jun* helices. Representative fragments with contiguous break point termini at Glu197 and Leu198 were designated  $\alpha$ 197 (N-terminal fragment) and  $\omega$ 198 (C-terminal fragment), and subsequently shown to produce selectable activity in the *E. coli* periplasm with interactions between a variety of heterologous domains fused to the break-point termini, including single-chain antibody Fv fragments (scFv), antibody light chains (LC), thioredoxin with 12-mer peptides inserted into the active site (trxpeps), and the extra-cellular domain of the B-cell activation antigen CD40 (CD40ED). Activation by complementation of  $\alpha$ 197 and  $\omega$ 198 could also be driven by interaction of the heterologous domains with a third polypeptide, such as a receptor. Contiguous break-point termini of interest in *E. coli* TEM-1  $\beta$ -lactamase in addition to E197/L198 include amide-bond junctions between amino acid residues N52/S53, E63/E64, Q99/N100, P174/N175, K215/V216, A227/G228, and G253/K254. The combined lengths of the fragment pairs may be discontinuous or overlapping, however, comprising from 90% to 110% of the total length of the parent protein, and the actual break-point could be within ten amino acid residues in either direction from an identified functional contiguous break-point junction. The specific activity of the reconstituted enzyme can be enhanced to near wild-type levels by the interaction-driven formation of a disulfide at the break-point, which restores the integrity of the native polypeptide backbone (see Figure 4).

The  $\beta$ -lactamase  $\alpha$ 197 and  $\omega$ 198 fragments cooperatively produce selectable activity

in the bacterial periplasm in a manner that is strictly dependent on specific interaction between heterologous domains fused to the break-point termini of the fragments is an example of an enzyme-based molecular interaction sensor that can undergo secretory translocation across a plasma membrane into an extra-cellular compartment, and therefore can reliably detect interactions between and among extra-cellular proteins.

The interaction-dependent enzyme association systems of the present invention finds use in many applications in human therapeutics, diagnostics, and prognostics, as well as in high-throughput screening systems for the discovery and validation of pharmaceutical targets and drugs.

One particular application is concerned with the localized and controlled activation of inactive or weakly active compounds. For example, many useful compounds, such as drugs, chromophores, and fluorophores, can be inactivated by conjugation of an essential moiety on the compound, such as a hydroxyl or amino group, to a substrate for enzymatic hydrolysis, such as an ester, amide, carbamate, phosphate, glycoside, or glucuronide (Jungheim and Shepherd, *Chem Revs.* (1994) 94:1553). Such conjugates can then be reactivated by the appropriate hydrolytic enzymes such as esterases, carboxypeptidases, alkaline phosphatases, glycosidases, glucuronidases,  $\beta$ -lactamases, and Penicillin-amidases.

In one particularly versatile system, cephalosporins may be conjugated at the 3' position via a variety of different leaving groups to a variety of anti-cancer drugs, such as nitrogen mustards, methotrexate, anthracyclines, and vinca alkaloids (Svensson *et al.*, *J Med Chem* (1998) 41:1507; Vrudhula *et al.*, *J Med Chem* (1995) 38:1380; Jungheim and Shepherd, 1994, *supra*; Alexander *et al.* *Tetrahedron Lett* (1991) 32:3269; see also Figure 5). All of these are good substrates for broad spectrum  $\beta$ -lactamases, and most are much less active than their parent drugs. As a result, these prodrugs are promising candidates for use in Antibody-Directed Enzyme Prodrug Therapy (ADEPT; Bagshawe, *Drug Devel Res* (1995) 34:220). In addition to these compounds a vast array of antibiotics (Holbrook and Lowy, *Cancer Invest* (1998) 16:405), as well as a variety of chromogenic and fluorogenic substrates have been developed for  $\beta$ -lactamases (Jones *et al.*, *J Clin Microbiol* (1982) 15:677; Jones *et al.*, *J Clin Microbiol* (1982) 15:954; Zlokarnik *et al.*, *Science* (1998) 279:84), making them one of the most versatile known classes of enzymes.

Nevertheless, the utility of such enzymes would be greatly enhanced if they could be engineered so that their catalytic activities could be positively controlled by allosteric

interaction with ligands of choice. In this way the catalytic power of these enzymes could be harnessed to multiple new applications, including (1) rapid, ultra-sensitive detection of trace analytes and pathogens in biological specimens or in food, (2) targeted activation of therapeutic and diagnostic reagents at specific locations in the body, (3) rapid enrichment of expressed sequence libraries for autonomously folding domains (AFDs), (4) massive parallel mapping of pair-wise protein-protein interactions within and between the proteomes of cells, tissues, and pathogenic organisms, (5) rapid selection of antibody fragments or other binding proteins to whole proteomes, (6) rapid antigen identification for anti-cell and anti-tissue antibodies, (7) rapid epitope identification for antibodies, (8) high-throughput screens for inhibitors of any protein-protein interaction.

For example, enzymes which could be activated to hydrolyze chromogenic substrates only upon binding to target analytes could form the basis of assays for those analytes of unparalleled sensitivity and convenience. Such assays would be homogeneous, requiring no manipulations other than the mixing of two components, namely the enzyme and substrate, with a biological specimen, in which the presence of the analyte is then quantitatively indicated by the rapid development of color. Current homogeneous enzymatic assays rely on inhibition of the enzyme by binding of anti-analyte antibody to the analyte, or mimic thereof, immobilized on the surface of the enzyme (Coty *et al.*, *J Clin Immunoassay* (1994) 17:144; Legendre *et al.*, *Nature Biotech* (1999) 17:67). Free analyte is estimated by its ability to competitively displace the antibody, thereby activating the enzyme. Such enzymes are thus activated competitively, not allosterically. For assays employing such enzymes the maximum signal increment occurs at equilibrium with roughly  $K_d$  concentrations of reagents, so that typically only a fraction of analyte molecules participates in signal generation, and equilibration is often slow or does not even reach completion. However, an enzyme which is activated by direct allosteric interaction with analyte, can be used in excess, so that equilibration is rapid and independent of the analyte concentration, and the analyte can be saturated to produce signal from every molecule. In the case of microbial or viral pathogens, where unique surface markers may be present in hundreds to thousands of copies per cell or particle, such enzymes, which would be activated by binding to the marker, could allow rapid detection of as little as a single cell or particle, whereas the sensitivity of equilibrium assays for such analytes would typically be much lower.

In another class of applications interaction-activated enzymes could be adapted for activation by binding to specific cell surface molecules. This would allow the enzyme to become localized and activated at specific sites in the body for target-restricted activation of reagents for therapy or imaging. Antibody-Directed Enzyme Prodrug Therapy (ADEPT; Bagshawe, 1995, *supra*) is a promising chemotherapeutic strategy for the treatment of cancer, in which a prodrug-activating enzyme, such as a  $\beta$ -lactamase, is targeted to the tumor by a tumor-specific antibody to which it is chemically or genetically conjugated. After unbound conjugate has cleared the circulation, an inactive prodrug, such as an anthracycline cephalosporin, is administered, which is converted to a potent tumor-killing cytotoxin at the site of the tumor by the remaining tumor-bound enzyme. The main problem with ADEPT is that the unbound conjugate must clear the circulation before the prodrug can be administered in order to minimize systemic toxicity. However, by the time the conjugate has cleared the circulation >90% of the tumor bound enzyme has been lost (Bagshawe, 1995, *supra*; Springer and Niculescu-Duvaz, *Anti-Cancer Drug Design* (1995) 10:361). In spite of this, ADEPT has been able to achieve higher active drug concentrations in the tumor than any other procedure (Sedlacek *et al.*, 1992 In *Contributions to Oncology*, Huber H and Queisser V, eds. pp. 208ff Karger, Basel), and has shown promise in the clinic (Bagshawe *et al.*, *Dis Markers* (1991) 9:233; Springer and Niculescu-Duvaz, 1995, *supra*; Martin *et al.*, *Cancer Chemother Pharmacol* (1997) 40:189). The unbound conjugate problem could be completely obviated by a prodrug-activating enzyme which would be active only when bound to the tumor, so that the prodrug could be administered simultaneously with the enzyme or at the point of peak tumor loading without regard for unbound enzyme which would be inactive.

In the same way, interaction-activated enzymes could be targeted for activation by surface markers on the cells of other types of diseased tissues, such as sites of inflammation or atherogenesis, or even healthy tissues. The target-localized and activated enzymes could then be used to activate not just cytotoxins, but other types of therapeutic agents such as small molecule agonists or antagonists of biological response modifiers, as well as imaging reagents for precise localization of tissue with disease or other phenotype of interest. For example, target-activatable enzymes could be used to deliver: (1) immune stimulants to tumors, (2) immuno-suppressants to sites of chronic inflammation or to organ transplants, (3) antibiotics to specific pathogens, (4) cytotoxins and anti-virals to virus-infected cells,

(5) hormones and other pleiotropic agents to specific cells and/or tissues, or (6) neurotransmitters and other neuro-modulators to specific nerves or tissues. In short, interaction-activated enzymes could be used to deliver to any tissue any small molecule cytotoxin, hormone, steroid, prostaglandin, neurotransmitter, or agonist/antagonist of peptide hormone, cytokine, or chemokine, etc., which could be inactivated by conjugation to the appropriate substrate.

In yet another class of applications, interaction-activated enzymes could be adapted for efficient simultaneous detection of multitudes of interactions among proteins within cells, including expressed sequence libraries, single-chain antibody fragment (scFv) libraries, and scaffolded peptide libraries. For example, enzyme-based interaction traps could enable the comprehensive mapping of pairwise protein-protein interactions within and between the proteomes of human cells, tissues, and pathogens for the rapid identification and validation of new pharmaceutical targets. They could also be used for rapid selection of binding molecules from single-chain antibody fragment (scFv) libraries, or from scaffolded peptide libraries for use as reagents in functional genomics studies, or for identification of natural ligands and epitopes by homology. Target interactions identified using interaction-dependent  $\beta$ -lactamases could be used immediately to screen for inhibitors of the interaction by exploiting the great substrate diversity of these enzymes to reverse the polarity of selection. Whereas interaction-dependent activation of  $\beta$ -lactamase could be used to confer selective growth on host cells in the presence of  $\beta$ -lactam antibiotics, it could also be used to confer selective cytotoxicity on the cells in the presence of  $\beta$ -lactam pro-antibiotics. The latter substrates would only become cytotoxic upon hydrolysis of the  $\beta$ -lactam moiety by the interaction-activated enzyme, and so could be used to select inhibitors of the interaction by their ability to confer selective growth on host cells.

Finally, enzyme-based interaction sensors could be used for rapid detection of the activation or inhibition of key molecular interactions in signal transduction pathways, enabling high-throughput cellular screens for inhibitors or activators of those pathways. For example, screening for agonists or antagonists of receptor tyrosine kinases usually requires coupling receptor ligation to a selectable phenotype which results from de novo gene expression. Such multi-step signal generating mechanisms are prone to high rates of false positive and false negative selection, like the yeast two-hybrid system, and are therefore poorly suited to high-throughput screening. However, interaction-dependent

β-lactamases could be set up for activation by phospho-tyrosine sensitive interactions, so that a selectable phenotype would be generated just downstream from receptor ligation. Interaction between the receptor tyrosine kinase substrate and a binder peptide could be designed to be either dependent on, or inhibited by phosphorylation, so that either receptor agonists or receptor antagonists could be selected.

### General Strategies for Making High-Performance Enzyme Fragment Complementation Systems

The present invention provides for general strategies for the use of heterologous interactors, break-point disulfides, random tri-peptide libraries, and mutagenesis to obtain stable enzyme fragments which are capable of forming of catalytically robust complexes. It has been suggested that it might be possible to identify such fragment pairs for any enzyme simply by conducting thorough searches of all possible fragment pairs for the enzymes in question. (Ostermeier *et al.*, *Proc Natl Acad Sci* (1999) 96:3562). In practice, however, the success of such endeavors is strongly dependent on the stringency of selection, that is, how much functional enzyme must be produced by the expressed fragments to produce an efficiently selectable phenotype. An efficiently selectable phenotype is one in which the background frequency, or false positive rate, is not appreciably higher than the frequencies of the desired fragments in the fragment libraries.

In fact the most useful fragment complementation systems for a given enzyme are not necessarily those fragments of wild-type sequence which are most capable of unassisted complementation, but rather the most useful fragment complementation systems comprise those fragments which, when using the engineering techniques described, can be made to meet more specific performance requirements. For example, naturally evolved proteins are generally expected to exhibit a roughly inverse correlation between fragment stability and complex stability. This is due to the energy cost of inter-conversion. The more stable the fragments are, the more energy is required to form the complex and vice versa. As a result, those fragments capable of producing the highest specific activities might be missed or dismissed because fragment instability may prevent them from producing selectable levels of activity. To circumvent such pitfalls, libraries of fragment pairs can be simultaneously expressed with libraries of random tri-peptides to insure that every fragment pair has a chance to perform in the presence of fragment-stabilizing tri-peptides, thereby

minimizing the dependence of the phenotype on fragment stability. This strategy is especially useful if dependence of activation on the interaction of heterologous domains fused to the fragments is desired. If constitutive activation is desired, the fragment libraries could also be amplified by error-prone PCR to introduce fold-accelerating mutations which could mitigate both fragment instability and complex instability, as was found for  $\beta$ -lactamase.

For *in vitro* applications such as homogeneous assays, biosensors, and target-activated reagents fragment stability is especially important, but the most stable fragments might not be selectable if they cannot produce stable complexes without assistance; as would be predicted by the inverse correlation of fragment stability and complex stability. Thus, fragment libraries could be expressed in the *E. coli* periplasm with a disulfide at the break-points and heterologous interactors fused to the break-point termini. These tools provide mechanisms for docking the fragments, accelerating folding, and stabilizing the active complex. As was shown with  $\beta$ -lactamase, a substantial fraction of fragment pairs can be made to produce robust selectable activity in the bacterial periplasm with such molecular prostheses.

Each of the four tools described for enhancement of functional reconstitution of the parent protein of the fragment pairs, i.e., heterologous interaction, break-point disulfide, tri-peptide stabilizers, and mutagenesis, can be used alone or in combination to insure selection of the best fragments for the desired application, and also to improve and optimize the performance of selected fragment pairs for a desired application. As demonstrated, each tool enhances performance by a different mechanism, so that the effects of multiple tools are generally additive. Heterologous interactors bring and hold the fragments together to facilitate re-folding into the active complex. Break-point disulfides can stabilize the active fold by restoring the integrity of the polypeptide backbone at the break-point. Tethered or free tri-peptides can protect the fragments from aggregation without interfering with folding into the active complex. Mutagenesis can protect the fragments by accelerating folding into the active complex.

The first step in the development of high-performance enzyme fragment complementation systems is to construct vectors to express each fragment in the fragment pair library. A convenient system for selective fragment library expression may be derived from the expression system illustrated in Figure 6. All fragment pairs regardless of the

intended application can potentially benefit from and would not be impaired by the docking function provided by interactors such as the fos and jun helixes fused to the break-point termini. Thus, the C-terminal, or  $\omega$  fragment library would be expressed as N-terminal fusions via a flexible polypeptide linker such as a (Gly<sub>4</sub>Ser)<sub>3</sub> linker to the fos helix (Interactor 2 in Figure 6) from the *lac* promoter in the phagemid vector pAO1 (the upstream cistron could be removed if desired). The amino acid sequence of the flexible polypeptide linker is not critical, however, it must be of a sufficient length and flexibility such that the fragment domain and heterologous interactor domain fold independently and unhindered. The N-terminal, or  $\alpha$  fragment library would be expressed as C-terminal fusions via a flexible polypeptide linker such as a (Gly<sub>4</sub>Ser)<sub>3</sub> linker to the jun helix (Interactor 1 in Figure 6) from the *trc* promoter in the compatible pAE1 vector. Coding sequences for signal peptides would be included if translocation to the periplasm were desired.

As discussed above, depending on whether the intended application(s) were *in vitro* or *in vivo*; or if *in vivo*, whether in the cytoplasm or secreted, one or more of the performance-enhancing tools may be incorporated into the expression vectors to maximize the probability of selecting the best fragment pair for the intended application(s). If periplasmic expression is desired, cysteines should be encoded at the break-point termini to allow disulfide formation. If the enzyme contains other cysteines, at least 1 mM and not more than 5 mM of a reducing agent such as GSH or DTT should be included in the growth medium to inhibit the formation of mixed disulfides. If fragment stabilization is desired to increase the importance of specific activity in selection, a random or VRK tri-peptide library may be encoded in frame with each fragment fusion between the break-point terminus and the flexible polypeptide linker. If VRK libraries were used for each fragment in a 50-fragment pair library, every possible tri-peptide-fragment combination would be contained in a combined library of  $<10^8$ . Alternatively, a single tri-peptide library could be used for each fragment pair in trans, as was described above. The tri-peptide library would be fused operably in frame *via* the flexible polypeptide linker to the N-terminus of thioredoxin and expressed from the upstream cistron in the pAO1 phagemid vector (see Figure 6).

The second step in the development of high-performance enzyme fragment complementation systems is to construct an expression library of candidate enzyme



fragment pairs. Methods for generating libraries of random fragment pairs have been described (Ostermeier *et al.*, 1999, *supra*). However, such libraries are quite inefficient as the vast majority of fragment pairs will be dysfunctional. For combinatorial screening of fragment pair libraries with mutagenic or random tri-peptide libraries, much more efficient fragment pair libraries will be necessary. For a variety of reasons it may be assumed that the most functional fragment pairs will correspond to scission of the polypeptide chain in exposed regions between elements of secondary structure. Exposed break-points will be required for use of tethered heterologous interactors and tri-peptides, and scission within secondary structure elements can irreversibly destabilize such elements. If a 3-dimensional structure is available for the enzyme of interest, or for a homolog, it can be used to identify exposed loops as candidate sites for chain scission. Typical globular proteins will not have more than 20-25 such sites that are far enough from the ends so that the larger fragment is not independently active. This is a manageable number for construction of coding sequences for each fragment pair by PCR. Two end-specific primers would be required, plus a head-to-head pair of primers for each break-point, which should be located more or less in the center of the exposed loop. If a 3-d structure is not available, reliable algorithms are available on the internet for computational prediction of secondary structure and hydropathy, such as the ProteinPredict program of Rost and Sander (*J Mol Biol* (1993) 232:584; *Proteins* (1994) 19:55; *Proteins* (1994) 20:216). With such programs, most of the exposed loops can be identified as hydrophilic regions between secondary structure elements. Again, it would not be excessively burdensome to prepare coding sequences by PCR for up to 50 fragment pairs.

If fragment complementation does not need to be dependent on the direct or ligand-mediated interaction of heterologous domains fused to the break-point termini, then fold-accelerating mutations could also be selected by using error-prone PCR in the initial amplification of the fragment coding sequences. Under appropriate conditions of  $Mg^{++}$ ,  $Mn^{++}$ , and nucleoside triphosphate concentrations, as well as cycle number, mutagenesis can be limited to 1-3 unbiased coding changes per molecule (Cadwell and Joyce, 1995, in *PCR Primer-A Laboratory Manual* C. Dieffenbach and G. Dveksler, Eds. Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 583-590). Since most mutations would be non-phenotypic, this could easily be combined with the other performance-enhancing tools without compromising the selectability of optimal fragment-tri-peptide combinations. Once

the fragment coding sequences have been amplified, gel-purified, and ligated into the vectors, the ligation products may be desalted and concentrated to allow efficient co-transformation of *E. coli* cells by high-voltage electroporation. If both the tri-peptide libraries and mutagenesis are used it is advisable to collect at least  $10^8$  and preferably at least  $10^9$  transformants to insure comprehensive representation of the full diversity of the library. The full library is then plated onto each of a range of non-permissive conditions, the least stringent being that on which the host cells would plate with an efficiency not greater than ten times the inverse of the library size. This would insure a manageable frequency of true positives among false positives. The maximum selection stringency would be that above which nothing is recovered from the library.

If fragment complementation is to be dependent on the direct or ligand-mediated interaction of heterologous domains fused to the break-point termini, then mutagenesis should not be used because folding acceleration usually eliminates the need for docking assistance. In this case selected fragment pairs must be counter-screened for loss of activity in the absence of the fos-jun interaction and activation indexes must be determined as the ratio of interaction-dependent activity to interaction-independent activity. For interaction mapping within or between proteome libraries activation indexes of the order of at least  $10^6$  are preferred since rare genes are expected to have frequencies in that range. For ligand-specific or interaction-specific biosensors lower activation indexes are usually acceptable. For example, to detect nanomolar concentrations of a ligand for which fragment-binder fusion affinities ( $K_d$ ) are in the 10 nM range, the fragment binder fusions need only to be used at 100 nM concentrations to saturate the ligand. Under these conditions ~90% of the fragment-binder fusions will be unbound. If the activation index is  $> 100$ , the background will be  $< 10\%$  of the signal.

Selected fragment pairs can be optimized for maximum activity and/or maximum activation index. In our experience break-point disulfides produce the highest specific activities because they allow the greatest amount of native structure in the fragment complex. However, they also may be in the background so that activation indexes are often lower. To retain the specific activity benefit of the break-point disulfide and reduce the background it may be necessary to retard the rate of disulfide formation so that it would not have sufficient time to occur during the abortive attempts of the unaided fragments to fold, but would occur efficiently when folding is catalyzed by the heterologous interaction.

Two parameters may be adjusted to control the formation of break-point disulfides. (1) The proximity of the disulfide-forming cysteines to the break-point may be adjusted to place greater orientational stringency on disulfide formation. (2) The concentration of reducing agent in the medium may be increased to reduce the effective concentration of DsbA, the principle disulfide-forming oxidase in the periplasm.

It is possible to use TEM-1  $\beta$ -lactamase fragment complementation to select fragment pairs of other proteins which do not produce selectable phenotypes in *E. coli* for their ability to form stable complexes because such complexes will usually be in the native conformation and should be functionally active. It has been amply demonstrated that naturally evolved proteins have unique minimum energy conformations in which they are stable and active (Li *et al.*, *Science* (1996) 273:666). All other conformations are unstable. Thus, if a fragment pair library of a non-phenotypic protein is expressed as fusions to the interaction-dependent TEM-1  $\beta$ -lactamase fragments, it is expected that only those fragment pairs which associate and fold into the native conformation will provide sufficient docking function to facilitate selectable  $\beta$ -lactamase activation. In this case, the subject fragments serve the purpose of the heterologous interactors in facilitating complementation of  $\beta$ -lactamase fragments. However, additional modifications could be encoded into the fragment/heterologous interactor fusion sequences to enhance functional reassociation of the  $\beta$ -lactamase fragments, including a break-point disulfide, a randomly-encoded peptide of from 3-12 amino acids, and mutagenesis of several amino acids within the fragment domain. All of these tools would specifically impact only complementation of the subject fragments by stabilizing the fragments, accelerating folding, and/or stabilizing the active fragment complex. Selected fragment pairs could then be tested individually for reconstitution of enzymatic activity or other function of the parental protein. In this way many useful fragment complementation systems could be developed for proteins which are active in eukaryotic cells, such as kinases or herbicide-resistance proteins.

The interaction-activated enzyme association systems of the subject invention, as exemplified by prokaryotic  $\beta$ -lactamase, find use in many applications as summarized below.

- (1) Simplex and multiplex protein-protein interaction mapping. Simplex refers to the use of single bait proteins to fish natural interactors out of expressed sequence libraries. Multiplex refers to the combinatorial pair-wise interaction of two expressed sequence

libraries for the purpose of simultaneously isolating as many natural interactions as possible. Individual interactors can be readily identified by nucleic acid hybridization.

- (2) Interaction-dependent  $\beta$ -lactamase systems may also be used to enrich randomly-primed expressed sequence libraries for fragments which encode autonomously-folding domains (AFD). Interference with folding by the fusion partner is avoided by using epitope tags and hetero-dimerizing helices only at the N- and C-termini of the expressed sequence, respectively. The fragments would have N- and C-terminal anti-tag binder and the partner hetero-dimerizing helix. The disulfide switch can accommodate diverse interaction geometries.

- (3) Simplex and multiplex selection of binding molecules such as single chain antibody fragments (scFv) and antibody light chain variable regions (VL). Non-immune human scFv repertoire libraries can be used with TEM-1  $\beta$ -lactamase interaction-dependent activation systems to isolate scFv to single baits or simultaneously to expressed sequence libraries. In the latter case scFv specific for individual targets can be readily identified by nucleic acid hybridization.

- (4) Interface mapping and ligand identification by mimotope homology. Constrained peptide libraries displayed on the surface of a carrier or "scaffold" protein may be used with  $\beta$ -lactamase interaction-dependent activation systems to isolate surrogate ligands for proteins or AFDs of interest. Consensus sequences from panels of such surrogate ligands for a given polypeptide may then be used to identify natural ligands of the polypeptide or interaction surfaces on natural ligands of the polypeptide. A common application of interface mapping is epitope mapping for antibodies, whereby the specific region to which an antibody binds on the surface of its antigen is identified.

- (5) Bio-Action Sensors. The efficiencies of most screening systems for signal transduction agonists and antagonists are compromised by the need for multiple steps between receptor ligation and selectable phenotype generation, which usually requires de novo gene expression. Interaction-activated  $\beta$ -lactamases can be tailored for activation or inhibition by any component of a target signal transduction pathway to allow selection of agonists or antagonists of the pathway in any appropriate cell type without the need to wait for gene expression to generate a selectable phenotype.
- (6) Homogeneous Assays. Interaction-dependent complementing fragments can be fused

to two scFv or other binding molecules which bind non-overlapping epitopes on target molecules, so that  $\beta$ -lactamase activation becomes dependent on binding to the target ligand. The use of ligand-dependent  $\beta$ -lactamases in homogeneous assays for two-epitope analytes from proteins to pathogens affords unparalleled sensitivity because saturation kinetics can be used instead of the equilibrium kinetics required by most assays. The binding molecules could also be oligonucleotides which anneal to contiguous sequences in the genome of a target pathogen. Such sequence-activated  $\beta$ -lactamases could also be used for rapid quantitation of specific PCR products without the need for gel electrophoresis.

(7) Target-Activated Enzyme Prodrug Therapy (TAcEPT) and Target-Activated Enzyme Imaging (TAcEI). Antibody-directed enzyme prodrug therapy is a promising chemotherapeutic strategy in which patients are treated with prodrug-activating enzymes such as  $\beta$ -lactamase conjugated to tumor-targeting antibodies (Bagshawe, 1995, *supra*). When unbound antibody-enzyme conjugate has cleared the circulation, prodrugs can be administered which are preferentially activated at the site of the tumor. The efficacy of this therapy is severely limited by the need for unbound conjugate to clear the circulation before the prodrug can be administered in order to avoid excessive toxicity, during which time most of the bound enzyme is lost from the tumor. The use of tumor-activated  $\beta$ -lactamases allows the prodrug to be administered at peak tumor loading of the enzyme since the latter is inactive in the circulation, and can only activate the prodrug when bound to the tumor. The same strategy can be used for antibody-directed site-specific activation of reagents for imaging of tumors or other tissue pathologies, or for other therapeutic indications such as inflammation or transplant rejection.

The following examples are offered by way of illustration of the present invention, not limitation.

## EXPERIMENTAL

### EXAMPLE 1

#### $\beta$ -lactamase Activation by Interaction-Mediated Complementation of $\alpha$ 197 and $\omega$ 198:

##### Interactions between scFv and trxpeps

5 This example demonstrates the ability of the system to detect and discriminate specific interactions between single chain antibody Fv fragments (scFv) and 12-amino acid peptides by inserted into the active site of *E. coli* thioredoxin (trxpeps, Colas *et al.*, *Nature* (1996) 380:548). ScFv are comprised of antibody heavy chain and light chain variable regions (VH and VL) tethered into a continuous polypeptide by most commonly a (Gly<sub>4</sub>Ser)<sub>3</sub> linker encoded between most commonly the C-terminus of VH and the N-terminus of VL.

10 ScFv from a human non-immune antibody repertoire were amplified by PCR using a consensus primer mix (Marks *et al.*, *Eur. J. Immunol.* (1991) 21:985); and subcloned into a pUC119-based phagemid vector (Sambrook *et al.*, *supra*) for expression of the scFv as fusions to the N-terminus of the  $\omega$ 198 fragment with an intervening (Gly<sub>4</sub>Ser)<sub>3</sub> linker (pAO1; see Figure 6A). An N-terminal signal peptide was provided for translocation to the bacterial periplasm. A commercial trxpep library was obtained and amplified by PCR using primers specific for the N- and C-termini of *E. coli* thioredoxin (Genbank accession no. M54881). This product was subcloned into a p15A replicon (Rose, *Nuc Acids Res* (1988) 16:355) for expression as fusions to the C-terminus of the  $\alpha$ 197 fragment from the *trp-lac* fusion promoter (pAE1; see Figure 6B). Again, an N-terminal signal peptide was provided for translocation to the periplasm. Figure 7 illustrates the activation of TEM-1 by complementation of  $\alpha$ 197 and  $\omega$ 198, mediated by interaction between an scFv and a 25 trxpep.

30 It was estimated that about 20% of the original scFv library clones produced soluble, full-length scFv as judged by immunoblot analysis (Harlow and Lane, (1988) *In Antibodies: A Laboratory Manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor) of periplasmic extracts obtained by osmotic shock (Neu and Heppel, *J Biol Chem* (1965) 240:3685). Thus, approximately 60 clones had to be screened in this way to obtain twelve clones expressing functional scFv. Plasmid DNA representing these twelve clones of the scFv- $\omega$ 198 construct was co-transformed with DNA representing approximately

5x10<sup>6</sup> clones of the  $\alpha$ 197-trxpep construct into *E. coli* strains DH5 $\alpha$  and TG1 (Sambrook *et al.*, 1989, *supra*), and plated onto solid LB medium containing kanamycin and chloramphenicol to determine the total number of co-transformants. Aliquots were also plated onto 25  $\mu$ g/ml ampicillin (amp25). Out of approximately 1x10<sup>7</sup> total co-transformants, 40 ampicillin-resistant clones were recovered, 36 of which replated on amp25. A similar number of co-transformants of a single randomly selected  $\alpha$ 197-trxpep construct with the twenty scFv- $\omega$ 198 constructs produced no colonies on amp25. All twelve scFv were represented in the 36 ampicillin-resistant clones with from one to five different trxpeps each. None of the 12 scFv cross-reacted with any trxpep originally selected by another scFv, as determined by co-transforming each scFv- $\omega$ 198 construct with a pool of the  $\alpha$ 197-trxpep constructs selected by the other scFv. Thus, all 36 selected clones were bona fide positives, representing unique and specific scFv-trxpep interactions. No scFv bound thioredoxin in the absence of its peptide mimotope(s), and no selected trxpep bound common determinants on the scFvs. Selections were performed in the *Es. coli* host strain TG1 without the gratuitous de-repressor of the *lac* promoter, isopropyl thiogalactoside (IPTG), so that transcription was minimal. When transcription was increased by the presence of 1 mM IPTG, many more colonies were obtained. Several of these were shown to be bona fide interactions which were too weak to confer selectable ampicillin resistance at lower levels of expression. Thus, the stringency of selection can be tuned by adjusting the expression levels of the interactors.

These results have several important implications. First, the false positive rate was exceedingly low, much lower than has been reported for other intra-cellular interaction sensors such as the yeast two-hybrid system (Bartel *et al.*, 1993, *supra*; Bartel *et al.*, 1996, *supra*). This property is essential for high-throughput applications. Secondly, the false negative rate with respect to the scFv was immeasurably low, as trxpeps were recovered for all functional scFv, and this too is essential for high-throughput applications. The fact that mimotopes were recovered for all scFv enables the system for high-throughput multiplex epitope mapping for scFv. Finally, the system is capable of efficient recovery of multiple interactions between two diverse populations of proteins simultaneously. Ultimately, given the high efficiency of the system, i.e., low rates of false positive and false negative selection, the throughput of the system should be limited only by the sizes of the interacting libraries, and/or the number of co-transformants which can be handled

conveniently. For example, construction of recombinant protein libraries in the  $10^9$ - $10^{10}$  range is routinely possible for scFv, trxpeps, or cDNAs (Hoogenboom *et al.*, *Immunotech* (1998) 4:1). Combinatorial pair-wise interaction trapping for any two such libraries would require at least  $10^{18}$ - $10^{20}$  clones, but with quantitative phagemid infection methods  
5 (Sambrook *et al.*, 1989, *supra*) and automated fermentation and plating methods, such throughput levels could be realistically achieved.

### Example 2

#### $\beta$ -lactamase Activation by Interaction-Mediated Complementation of $\alpha$ 197 and $\omega$ 198:

##### Interactions between antibody light chain V-regions (VL) and trxpeps

This example demonstrates the ability of the system to work with larger antibody fragments, such as Fab, which are comprised of entire light chains disulfide-bonded to Fd fragments which contain VL plus the first heavy chain constant region. A subset of Fabs from a human repertoire library was subcloned for expression as C-terminal  $\omega$ 198 fusions  
10 from a dicistronic transcript from the *lac* promoter in the pAO1 vector (see Figure 6A). The first cistron encoded the light chain with a signal peptide for translocation to the periplasm. The light chain termination codon was followed by a short spacer sequence and then a ribosome binding site approximately 10 bp upstream from the start of translation for the signal peptide of the Fd fragment, which was followed by  $\omega$ 198 with an intervening  
15 (Gly<sub>4</sub>Ser)<sub>3</sub> linker. This construct was then co-expressed with the  $\alpha$ 197-trxpep library in the pAE1 vector in strains DH5 $\alpha$  and TCG1. Spontaneous association of the light chain with the Fd- $\omega$ 198 fusion protein in the periplasm was expected to produce a functional Fab  
20 fragment. Binding of the latter to the peptide on a  $\alpha$ 197-trxpep fusion was then expected to facilitate assembly of the functional TEM-1  $\beta$ -lactamase in amounts sufficient to confer  
25 selectable resistance to ampicillin on the host cells.

Many clones were in fact recovered on 25  $\mu$ g/ml ampicillin. Some of these are listed in Table 1 below. Several were resistant to up to 100  $\mu$ g/ml and one was resistant to up to 600  $\mu$ g/ml. Unexpectedly, all recovered Fabs were missing the VH region. That is, they contained the full-length light chain (LC) with only the first heavy chain constant  
30 region (CH1). The reasons for this were as follows. The original Fab library was constructed by first inserting the VL repertoire into the vector which already contained the constant regions ready for expression. This intermediate construct was capable of



expressing a complex of the light chain with the first heavy chain constant region fused to  $\omega$ 198. Plasmid DNA was then purified from this light chain library and used as the recipient for insertion of the VH repertoire to complete the Fab library. The resulting library was contaminated with approximately 15% of clones which contained the intermediate vector. Only these LC-CH1 complexes were capable of driving  $\alpha$ 197- $\omega$ 198 complementation by binding of the VL combining site with the peptide on the appropriate trypsin. It is not known why full-length Fabs were not selected, however, the larger size and rigidity of the Fab-trypsin complex (~67 kDa) may have sterically inhibited fragment complementation, whereas the smaller size and flexibility of the LC-CH1 complex did not.

TABLE 1.

**Ampicillin-Resistance of TEM-1  $\beta$ -lactamase  $\alpha$ 197/ $\omega$ 198 Fragment  
Complementation Driven by Interaction of Selected Pairs of  
Antibody Light Chain-CH1 Complexes and Trypsins**

<u>LC-CH1</u>	<u>Trypsin</u>	<u>Amp<sup>r</sup></u>
P44-2-2B1	P44-2-2A1	+++++ <sup>a</sup>
P44-2-3B1	P44-2-3A1	++
P44-1-6B1	P44-2-6A1	+
P64-17B1	P64-17A1	++
P65-1-10B1	P65-1-10A1	+++
P66-3-2B1	P66-3-2A1	++
P66-3-10B1	P66-3-10A1	+
P66-3-14B1	P66-3-14A1	++
P75-7-7	?	$\geq$ +
P75-7-13	?	$\geq$ +
P75-7-30	?	$\geq$ +

<sup>a</sup> +, ++, +++, +++++, >10% plating efficiency on 25, 50, 100, 600  $\mu$ g/ml ampicillin.

This result shows that light chain V-regions alone, which are only ~12 kDa in size, could make convenient high-affinity binding molecules for antigen-dependent activation of  $\beta$ -lactamase by fragment complementation. To test this, the VLs from several of the selected LC-CH1 were subcloned for expression alone as C-terminal fusions to  $\omega$ 198.

When each was co-expressed with its partner  $\alpha$ 197-trxpep, approximately one-third of the VL conferred selectable resistance to ampicillin comparable to the parent LC-CH1s.

### Example 3

#### $\beta$ -lactamase Activation by Interaction-Mediated Complementation of $\alpha$ 197 and $\omega$ 198:

##### Interactions between CD40 and trxpeps

This example demonstrates the ability of the present system to isolate panels of trxpeps that bind to a given protein of interest, and which could be used to map interaction surfaces on the protein, and which could also assist in the identification of new ligands by homology. The extra-cellular domain of the human B-cell activation antigen CD40 is known to reliably express in the *E. coli* periplasm (Noelle *et al.*, *Immunol Today* (1992) 13:431; Bajorath and Aruffo, *Proteins: Struct, Funct, Genet* (1997) 27:59). A T-cell surface molecule, CD40 ligand (CD40L), is known to co-activate B-cells by ligation to CD40, but there may be other ligands. Therefore, TEM-1  $\alpha$ 197/ $\omega$ 198 fragment complementation was used to select a panel of CD40-binding trxpeps. The sequences of these peptides would then be examined for homology to the known ligand and other potential ligands. The coding sequence for the mature form of the extra-cellular domain (CD40ED) was amplified by PCR using primers homologous to the N-terminus of the mature protein and to the C-terminus of the ~190-residue extra-cellular domain (Genbank accession no. X60592). The PCR product was then subcloned into the pAO1 phagemid vector (Figure 6A) for expression from the *lac* promoter as a C-terminal fusion to the TEM-1  $\omega$ 198 fragment with an intervening (Gly<sub>4</sub>Ser)<sub>3</sub> linker. Expression of the correct product was confirmed by PAGE, and the CD40 fusion vector was then rescued as phage and transfected into TG-1 cells bearing the same trxpep library construct as described above. Approximately 10<sup>7</sup> co-transformants were collected by double selection on kanamycin and chloramphenicol, and then plated onto 25  $\mu$ g/ml ampicillin. Activation of TEM-1 by a trxpep-CD40 interaction-mediated complementation of  $\alpha$ 197 and  $\omega$ 198 is depicted in Figure 8.

Ampicillin-resistant clones encoding thirteen unique trxpeps were recovered. In all cases amp resistance was strictly dependent on the presence of CD40ED and the peptide portion of the trxpep. No activity was seen if CD40ED was replaced with an irrelevant protein or if the trxpep was replaced by wild-type thioredoxin. The sequences of the

selected CD40-binding peptides are shown in Table 2 below along with their homologies to each other and to CD40L. The thirteen peptides sort into eight homology groups: two groups with three each (1 and 2), one with two (3), and five with one each. Groups 1 and 2 are defined by homology of three peptides in each group to the same region of CD40L.

5 Group 1 is homologous to the region of CD40L from Pro217 to Gly234, and Group 2 is homologous to the region from Gly158 to Leu168. Group 3 is defined only by inter-peptide homology and has no detectable homology to CD40L. Group 4 is homologous to CD40L from Ser110 to Pro120, and Group 5 is homologous to CD40L from Pro244 to Gly257. Groups 6-8 have no discernable homologies. However, a number of the peptides  
10 had striking homology to other human extra-cellular proteins, including CTLA-2A, a matrix metalloproteinase, a receptor Tyr phosphatase, vascular endothelial cell growth inhibitor (VEGI), transferrin receptor, CD3 $\zeta$ , and bone morphogenetic protein 3B (BMP-3B). These may define an interaction motif or motifs, which have been used repeatedly for extra-cellular protein-protein interactions. They may also indicate multiple interaction sites  
15 on CD40.

Inter-trxpep competition was tested by expressing each of five selected CD40-binding trxpeps from a second cistron in the pAO1 phagemid vector, downstream from the CD40 -  $\omega$ 198 fusion. Each of these constructs was then co-expressed with each of the same five plus three additional selected  $\alpha$ 197-trxpep fusion constructs in strain TG1 and  
20 scored for growth on 25  $\mu$ g/ml ampicillin. The results are shown in Table 3 below. The eight trxpeps sorted into five groups. BW10-1 competes moderately with groups 2 and 3. p58-12-9A1, BW10-4, and BW10-8 compete strongly with each other and have similar competition profiles. They do not compete with group 3, except for BW10-8, which competes slightly with group 3 and BW10-9. All three compete with BW10-1, and p58-12-  
25 9A1 also competes slightly with BW10-9. p44-4-2A1 and p45-7-2A3 compete strongly and have similar competition profiles. They compete with BW10-1 and nothing else except BW10-8 slightly. BW10-9 competes slightly with BW10-8 and p58-12-9A1. p65-2-9A1 is inhibited by nothing.

5/21

Table 2

### Homologies of Representative CD40-binding Trxpeps

Group	TrxPep	Sequence <sup>a</sup>	Amp <sup>r</sup>
1	BW10-1	<u>CGP</u> <u><b>K</b>ELRIGGRPRR</u> <u><b>G</b>P</u> C	<sup>b</sup> +
	P58-12-9A1	<u>CG</u> <u>PE</u> <u>G</u> <u>G</u> <u>G</u> <u>V</u> <u>A</u> <u>V</u> <u>G</u> <u>G</u> <u>V</u> <u>G</u> <u>G</u> <u>P</u> <u>C</u>	+
	P65-2-4A2	<u>CG</u> <u>PA</u> <u>K</u> <u>R</u> <u>A</u> <u>D</u> <u>V</u> <u>E</u> <u>F</u> <u>S</u> <u>L</u> <u>E</u> <u>P</u> <u>G</u>	+
	CD40L	215-AKPCGQQSIHLGGVFELQPGA-235	
2	BW10-9	<u>CG</u> <u>PK</u> <u>S</u> <u>A</u> <u>G</u> <u>K</u> <u>R</u> <u>K</u> <u>D</u> <u>R</u> <u>R</u> <u>K</u> <u>G</u> <u>P</u> <u>C</u>	++
	P65-2-1A3	<u>CG</u> <u>P</u> <u>R</u> <u>T</u> <u>R</u> <u>V</u> <u>N</u> <u>H</u> <u>Q</u> <u>G</u> <u>K</u> <u>T</u> <u>R</u> <u>G</u> <u>P</u> <u>C</u>	+
	P65-2-2A5	<u>CG</u> <u>PA</u> <u>G</u> <u>A</u> <u>I</u> <u>R</u> <u>H</u> <u>E</u> <u>H</u> <u>R</u> <u>Q</u> <u>L</u> <u>G</u> <u>P</u> <u>C</u>	+
	CD40L	152-LVTLENGKQLTVKRQGLYYIYAQ-174	
3	P44-4-2A1	<u>CG</u> <u>P</u> <u>D</u> <u>T</u> <u>G</u> <u>L</u> <u>E</u> <u>T</u> <u>D</u> <u>A</u> <u>A</u> <u>D</u> <u>A</u> <u>S</u> <u>G</u> <u>P</u> <u>C</u>	+
	P45-7-2A3	<u>CG</u> <u>P</u> <u>R</u> <u>R</u> <u>V</u> <u>R</u> <u>E</u> <u>T</u> <u>V</u> <u>A</u> <u>V</u> <u>E</u> <u>S</u> <u>S</u> <u>G</u> <u>P</u> <u>C</u>	+
4	BW10-4	<u>CG</u> <u>P</u> <u>P</u> <u>C</u> <u>A</u> <u>T</u> <u>F</u> <u>E</u> <u>E</u> <u>A</u> <u>K</u> <u>S</u> <u>N</u> <u>Q</u> <u>G</u> <u>P</u> <u>C</u>	+
	CD40L	104-ETKKENSFEMQKGDQNPQ-121	
5	P65-2-8A3	<u>CG</u> <u>P</u> <u>G</u> <u>R</u> <u>E</u> <u>S</u> <u>R</u> <u>G</u> <u>R</u> <u>C</u> <u>Y</u> <u>T</u> <u>P</u> <u>S</u> <u>G</u> <u>P</u> <u>C</u>	+
	CD40L	242-TDPSQVSHGTGFTSFGLL-259	
6	BW10-8	<u>CG</u> <u>P</u> <u>N</u> <u>T</u> <u>P</u> <u>D</u> <u>E</u> <u>E</u> <u>M</u> <u>A</u> <u>P</u> <u>Q</u> <u>A</u> <u>P</u> <u>G</u> <u>P</u> <u>C</u>	++
7	P65-2-5A4	<u>CG</u> <u>P</u> <u>V</u> <u>V</u> <u>H</u> <u>I</u> <u>K</u> <u>T</u> <u>N</u> <u>E</u> <u>Q</u> <u>A</u> <u>A</u> <u>P</u> <u>G</u> <u>P</u> <u>C</u>	+
8	P65-2-9A1	<u>CG</u> <u>P</u> <u>V</u> <u>A</u> <u>E</u> <u>E</u> <u>P</u> <u>A</u> <u>G</u> <u>G</u> <u>A</u> <u>R</u> <u>P</u> <u>G</u> <u>P</u> <u>C</u>	+

<sup>a</sup>For sequence homologies, underlined denotes identity, bold denotes conservative substitution. For groups 1, 2, 4, and 5 homologies to CD40L only are depicted.

4, and 5 homologies to CD40L only are depicted.

<sup>b</sup>Plating efficiencies for CD19c only are depicted.  
+ , > 10%; ++ , > 50%.

**Table 3**  
**CD40 Trx-Peptide Competition**

	<u>B10-1</u>	<u>B10-4</u>	<u>B10-8</u>	<u>B10-9</u>	<u>P44-4-2A1</u>	<u>P45-7-2A3</u>	<u>P58-12-9A1</u>	<u>P65-2-9A1</u>
B10-1	+	+/+	±/±	-/-	±/±	+/+	-/+	-/-
B10-4	+/+	+	+/+	-/-	-/±	-/-	+/+	-/-
B10-8	±/±	+/+	+	±/±	-/±	±/±	+/+	-/-
B10-9	-/-	-/-	±/±	(+)	-/-	+/+	+/+	-/-
P44-4-2A1	+/±	-/±	-/±	-/-	+	+/+	-/-	-/-
P45-7-2A3	+/+	-/-	±/±	+/+	+/+	(+)	+	-/-
P58-12-9A1	-/+	+/+	+/+	+/+	-/-	+	+	-/-
P65-2-9A1	-/-	-/-	-/-	-/-	-/-	-/-	-/-	(+)

Group 1: B10-1

Group 2: P58-12-9A1, B10-8, B10-4

Group 3: P44-4-2A1, P45-7-2A3

Group 4: B10-9

Group 5: P65-2-9A1

1. "+" = inhibited, "-" = not inhibited. Read down / across
2. For all cells right of "+" diagonal, read down=free / across=α-fusion.
3. For all cells left of "+" diagonal, read down=α-fusion / across=free
4. (+) self control was not actually done.

In general, the competition data is consistent with the homology data with the caveat that simultaneous binding to non-overlapping epitopes is sometimes not tolerated. This allows unrelated sequences like p58-12-9A1 and BW10-8 to compete strongly with one another and have similar competition profiles. This is probably due to steric interference with enzyme reassembly, and may account for the discordance between homology and competition data for BW10-1 and p58-12-9A1 in particular. These two probably bind near the same CD40 interaction epitope, which may sterically inhibit fragment complementation for many (but not all) other trxpeps.

For some applications it will be useful for  $\beta$ -lactamase activation to be mediated by simultaneous binding of both  $\alpha$ 197 and  $\omega$ 198 to non-overlapping epitopes on a separate molecule, either a free ligand or cell surface receptor. Two CD40-binding trxpeps, which had been identified as non-competing by the competition tests, were used to test this utility.

One of the two trxpeps was subcloned for expression as the C-terminal  $\omega$ 198 fusion from the pAO1 vector (see Figure 6). The other trxpep was expressed as the  $\alpha$ 197 fusion from the pAE1 vector as before. Co-expression of these two constructs was used as the negative control. To test for CD40-mediated activation, the CD40ED coding sequence (including signal peptide) was subcloned into the trxpep- $\omega$ 198 expression cassette between the promoter and the trxpep- $\omega$ 198 sequence. An additional 20 bp containing a ribosome binding site was included downstream from the CD40 stop codon to allow expression of both CD40 and trxpep- $\omega$ 198 from the same dicistronic transcript, as was described above for the Fab. As shown in Table 4 below, CD40 expression induced resistance to 50  $\mu$ g/ml ampicillin, whereas without CD40 the cells expressing the control constructs produced fewer than  $10^{-6}$  colonies per cell on 25  $\mu$ g/ml ampicillin. Thus,  $\beta$ -lactamase fragment complementation can be efficiently induced by a tri-molecular protein-protein-protein interaction.

Table 4

**Ligand activation of TEM-1 $\alpha$ / $\omega$  fragment complementation using non-competing CD40-binding trxspeps and CD40ED.**

<u>Molecule#1</u>	<u>Molecule#2</u>	<u>Molecule#3</u>	<u>Amp<sup>r</sup></u>
$\alpha$ -p44-4-2	CD40- $\omega$	-	++
$\alpha$ -p44-4-2	CD40	BW10-1- $\omega$	++
$\alpha$ -p44-4-2	-	BW10-1- $\omega$	-

a. plating efficiencies on 25  $\mu$ g/ml ampicillin in colonies per cell. -, <10<sup>-6</sup>;  
+, >10%; ++, >25% ++++, >50%.

#### Example 4

##### **$\beta$ -lactamase Activation by Interaction-Mediated Complementation of $\alpha$ 197 and $\omega$ 198:**

##### **Interaction between a CD40-specific scFv and CD40.**

Since  $\beta$ -lactamase activation by  $\alpha$ 197- $\omega$ 198 fragment complementation could be driven efficiently by interaction between scFv and trxspeps, it was important to show that it could also be driven by interaction between scFv and a *bona fide* protein antigen, preferably a cell surface receptor. This was especially important because the ligand-binding domains for type 1 trans-membrane receptors are N-terminal, therefore their expression as C-terminal fusions is preferred. However, the preferred orientation for scFv expression is also N-terminal. To allow expression of both scFv and antigen as C-terminal fusions,  $\beta$ -lactamase activation by a tri-molecular interaction was tested, including the C-terminal fusion of the scFv with  $\omega$ 198, a C-terminal fusion of CD40 with the fos helix, and a C-terminal fusion of  $\alpha$ 197 with the jun helix. The expression constructs were analogous to those used for CD40 ligation of the trxspep-fragment fusions. The CD40-fos fusion and the scFv $\omega$ 198 fusion were expressed from a dicistronic transcript in the pAO1 vector, and  $\alpha$ 197-jun fusion was expressed from the pAE1 vector. The fos-jun interaction has a  $K_d$  in the 10<sup>-8</sup>M range, so it should quantitatively ligate CD40 with  $\alpha$ 197, which are much more abundant than this in the periplasm. Binding of the scFv to CD40 should then dock  $\omega$ 198 with the complex to facilitate fragment complementation. As shown in Table 4, CD40-fos expression induced resistance to up to 100 $\mu$ g/ml ampicillin, whereas cells expressing only the control constructs without CD40-fos again produced fewer than 10<sup>-6</sup> colonies per cell

on 25  $\mu\text{g/ml}$  ampicillin. Thus,  $\beta$ -lactamase fragment complementation can be efficiently induced by a tri-molecular interaction of two extra-cellular proteins in preferred C-terminal fusions.

### Example 5

#### Disulfide-Enhanced Fragment Complementation

The  $\beta$ -lactamase activity produced by interaction-dependent complementation of the  $\alpha 197$  and  $\omega 198$  fragments is substantially less than that of the wild-type enzyme under the same expression conditions. This loss of activity could be due to a tendency of the fragments to aggregate or turnover when they are not folded into the native conformation, and it could also reflect a loss of specific activity due to the reduced ability of the loosely tethered heterologous interaction to stabilize the native conformation. It was reasoned that both folding kinetics and stability could be enhanced by the introduction of a disulfide at the break-point, and this could lead to a substantial increase in interaction-dependent activity. The expectation was that when the fragments were docked by the heterologous interaction, the integrity of the polypeptide backbone would be restored at some point in the folding pathway by the formation of a disulfide linkage between cysteines added at the break-point, and this would accelerate folding and/or stabilize the active conformation. The disulfide would form very rapidly in the highly oxidizing environment of the bacterial periplasm. However, if the fragments were unstable until they were docked and folded, but once folded the activity was stable, then the break-point disulfide might have little effect on activity if it did not form until late in the folding pathway.

Cysteines were added to the sequences of  $\alpha 197$  and  $\omega 198$ , between the break-point termini and the linkers leading to the heterologous interactors. With the fos and jun helices as the interactors, quantitative ampicillin resistance ( $> 10\%$  plating efficiency) increased from 50  $\mu\text{g/ml}$  to more than 100  $\mu\text{g/ml}$ , and the plating efficiency on 25  $\mu\text{g/ml}$  ampicillin increased at least 2-fold. Thus, disulfide formation must be accelerating folding and/or stabilizing the active conformation. However, the disulfide produced nearly as much activity without the interactors. This contrasts sharply with the activity of the fragments in the absence of either the disulfide or interactors, for which plating efficiencies are less than  $10^{-6}$  on 25  $\mu\text{g/ml}$  ampicillin. This result suggests that the fragments probably associate and refold readily on their own at these intra-cellular concentrations, but that



without a heterologous interaction or disulfide at the break-point, either folding cannot progress to the active conformation, or the latter is not stable enough to produce selectable activity. There must be a finite window of opportunity for disulfide formation when the thiols are proximal during unassisted folding. This window should be much wider during interaction-assisted folding. Thus, it should be possible to retard disulfide formation and thereby make it more dependent on the heterologous interaction.

Disulfide formation was made to be more dependent on the heterologous interaction by two modifications. First, disulfide formation could be inhibited by inclusion of a reducing agent in the growth medium. Dithiothreitol (DTT) at 10 mM reduced the plating efficiency of the disulfide-assisted fragments on 100 µg/ml ampicillin to  $< 10^4$  colonies per cell in the absence of an interaction, whereas with the fos-jun interaction the activity of the same fragments was little affected by DTT, so that the activation index was increased to  $> 1000$ -fold. Secondly, the cysteines were shifted by one residue each away from the break-point and into the  $\beta$ -lactamase sequence, so that they became separated in the native fold by an additional  $\sim 8\text{\AA}$ . This reduced activity to a plating efficiency of  $< 10^6$  on 50 µg/ml ampicillin without the interaction, whereas with the fos-jun interaction the plating efficiency was reduced to  $\sim 10\%$  on 50 µg/ml ampicillin for an activation index of  $> 10^5$ . Thus, a combination of reducing agent and thiol separation may be expected to increase the increment of interaction-dependent activation over background even further, perhaps to  $> 10^6$ . In any case the 8 Å increase in thiol separation alone increased the activation increment substantially over that of the fos-jun interaction without disulfide. The enhancement of interaction-dependent specific activity provided by the disulfide should allow weak interactions and/or poor expressors to produce selectable  $\beta$ -lactamase activity with fewer than 10 molecules per cell of the activated enzyme.

The ability of the break-point disulfide to enhance activation of TEM-1  $\alpha 197/198$  fragment complementation, suggests that break-point disulfides might be able to activate many enzyme fragment pairs which produce weak or no selectable activity with a heterologous interaction alone. The heterologous interaction may be essential for fragment docking, but since it is tethered with  $\sim 60\text{\AA}$  linkers it cannot restore the tight junction of the polypeptide backbone at the break-point. However, formation of a disulfide across the break-point should restore the integrity of the backbone, and should thereby help stabilize the active site of the complex. This idea was tested by screening nine additional pairs of

TEM-1  $\beta$ -lactamase fragments, corresponding to scission in nine exposed loops of the polypeptide chain. The nine fragment pairs were screened for selectable activity with the break-point disulfide alone, the fos-jun interaction alone, and with both together. The results are summarized in Table 5.

Addition of the break-point disulfide to the fos-jun interaction strongly increased the activity of seven of the nine fragment pairs, which makes eight out of ten pairs when  $\alpha 197/\omega 198$  is included. The ten fragment pairs may be sorted into three groups. One group comprises the two negative pairs. The second group comprises three pairs which can only be activated by disulfide and fos-jun interaction together. In each case, the plating efficiency is at least 10% on 25  $\mu\text{g/ml}$  ampicillin, with an activation index of at least 1000. The third group comprises five pairs, all from break-points in the C-terminal third of the molecule, which produce modest-to-robust activity with fos-jun alone, but no potent activity with both fos-jun and the disulfide together. Most importantly, four of these five produce no selectable activity with the disulfide alone, so they have very large activation indexes. P174/N175 had the highest activation index,  $\approx 10^7$  on 100  $\mu\text{g/ml}$  ampicillin. G253/K254 had the highest activity with a plating efficiency of  $>25\%$  on 400  $\mu\text{g/ml}$  ampicillin. Interestingly, the first fragment pair identified to exhibit interaction-dependent activation,  $\alpha 197/\omega 198$ , remains the only pair to produce robust selectable activity with the break-point disulfide alone. It is possible that activation of some pairs is inhibited by the formation of mixed disulfides between the break-point cysteines and the internal cysteines, and it is also possible that such inhibition could be alleviated with exogenous reducing agent. However, it is at least as likely that in these cases unassisted refolding could not proceed far enough to allow efficient formation of the break-point disulfide before aborting.

## Table 5

<u>Break-Point</u>	<u>+S-S, +Fos/Jun</u>		<u>+S-S</u>		<u>+Fos/Jun</u>		<u>Background</u>	
	<u>Amp25</u> <sup>b</sup>	<u>Hi Amp</u> <sup>c</sup>	<u>Amp25</u>	<u>Hi Amp</u>	<u>Amp25</u>	<u>Hi Amp</u>	<u>Amp25</u>	<u>Hi Amp</u>
N52/S53	+	25	—	—	—	—	—	—
E63/E64	+	25	—	—	—	—	—	—
L91/G92	—	—	—	—	—	—	—	—
Q99/N100	+	25	—	—	—	—	—	—
H158/V159	—	—	—	—	—	—	—	—
P174/N175	++++	200	—	—	+++	50	—	—
E197/L198	++++	100	+++	50	+++	50	—	—
K215/V216	++++	100	—	—	++	25	—	—
A227/G228	++++	200	—	—	+++	50	—	—
G253/K254	++++	400	—	—	+++	50	—	—

a. Fragment pairs were expressed in TG1 cells and plated onto ampicillin in the presence of 1mM IPTG. Fragments were expressed with or without break-point terminal thiols (S-S) and with or without break-point terminal fos ( $\omega$ ) or jun ( $\alpha$ ) helices.

<sup>b</sup>. Activities are expressed as plating efficiencies (colonies per cell) on 25 µg/ml ampicillin (amp25). -, <10<sup>-4</sup>; +/-, 0.01; +, 0.10; ++, 0.25; +++, 0.50; +++++, >0.90.

<sup>c</sup>. HiAmp refers to the maximum ampicillin concentration in  $\mu\text{g/ml}$  on which fragment-expressing cells plate with  $>10\%$  efficiency.

The fact that the fragment pairs which produced the highest activities are not the same as those with the highest activation indexes and vice versa, indicates that different fragment pairs may be optimally suited for different applications. For example, the activation index is more important than maximum activity for intra-cellular interaction mapping, where natural interactions must be identified against backgrounds of  $10^6$  or more non-interacting pairs. Thus, P174/N175 may be the best fragment pair for intra-cellular interaction mapping. On the other hand, maximum activity is more important than the activation index for in vitro applications because the activating target ligands will always be limiting in such applications. Since for maximum activation the fragments need only be used in ten-fold excess over their  $K_d$ s for the ligand, the activation index need only be 1000 for a signal-to-noise ratio of 100. Thus, G253/K254 may be the best fragment pair for in vitro applications such as biosensors or homogeneous assays.

The break-point disulfide overcomes a significant shortcoming of interaction-dependent enzyme fragment complementation systems: It is essential for high-throughput applications that such systems be capable of efficient activation by a wide range of heterologous protein-protein interactions. In other words, to minimize the false negative rate, the system must be activatable by any interaction between two proteins or fragments within the size range of single, naturally evolved protein domains, i.e., between ~100 and 300 amino acids in length. Globular proteins in this size range have radii in the range ~30-50Å. This means that the points of attachment for the linkers could be up to 100Å apart, and this distance must be spanned by the linkers in order for the break-points of the fragments to be able to come together. For this reason, the (Gly<sub>4</sub>Ser)<sub>3</sub> linker was selected, which is expected to be fully extended and flexible, and to have a length of ~60Å, thereby providing a combined length of up to 120Å to allow close approach of the break-point termini during folding. Nevertheless, it is reasonable to expect the stability of the active conformation to be quite sensitive, and generally inversely proportional to the dimensions of the heterologous interaction. Thus, for all such systems described to date it may be assumed that the longer the linkers, the larger the proportion of possible interactions that can accommodate refolding, but the less the interaction can contribute to stabilization of the active conformation.

The break-point disulfide overcomes this limitation because, if the linkers are long enough, it will form readily during re-folding, and once the break-point disulfide is formed

the specific activity of the reconstituted enzyme should be independent of the dimensions of the heterologous interaction, and in fact should not even require the continued integrity of the interaction. Thus, the break-point disulfide acts as a one-way switch, with an activation energy which can be supplied by a broad range of heterologous interactions, limited only by the ability of the interactors to fold properly, and by the length of the linkers to allow close approach of the break-point cysteines. This has two important consequences which allow a larger proportion of natural interactions to produce selectable activity. Longer linkers can be used, and interactions which are too weak to sustain selectable enzyme activity by themselves should still be able to "throw the disulfide switch" to produce selectable activity.

### Example 6

#### Peptide-Enhanced Fragment Complementation

Another way to enhance interaction-dependent enzyme fragment complementation is to introduce short, random peptide sequences at the break-points, and then to select for increased activity with a model interaction. Such peptide-dependent enhancements could occur by any of several mechanisms. For example, the peptides could stabilize the active conformation of the reconstituted enzyme by interacting with each other or with the enzyme itself, or the peptides could stabilize one or both of the fragments, thereby increasing steady-state activity by increasing fragment concentration.

Synthetic oligonucleotides were used to add three randomized residues to each fragment between the break-point residue and the linker for the heterologous domain. As the model interaction, the c-fos helix at the N-terminus of  $\omega$ 198 and the c-jun helix at the C-terminus of  $\alpha$ 197 was used. For each randomized position, a degenerate codon was used, which encoded a subset of amino acids which was biased toward charged residues to favor charge-charge interactions, which are the strongest. The VRK codon places c, a, or g in the first position, a or g in the second position, and t or g in the third position. The encoded amino acids are His, Gln, Arg, Asn, Lys, Ser, Asp, Glu, and Gly. For three randomized positions in both fragments there are a total of  $12^6 = 3 \times 10^6$  possible codon combinations, and  $9^6 = 5.3 \times 10^5$  possible different amino acid sequences. Initially, ten thousand clones of the library were plated onto successively higher concentrations of ampicillin until no colonies were recovered. Six clones in the DH5 $\alpha$  strain were recovered

from 800  $\mu\text{g/ml}$  ampicillin, and all six showed strict dependence on the fos-jun interaction for growth. In fact, the jun helix was removed from  $\alpha 197$  in the same starting  $10^4$  clones of the library, and when these clones were plated onto the same concentrations of ampicillin, only a few colonies grew on 200  $\mu\text{g/ml}$  ampicillin, and no colonies appeared on higher concentrations. This level of ampicillin resistance is comparable to that produced by the fos-jun interaction alone.

Unexpectedly, all six selected clones recovered from DH5 $\alpha$  had the same  $\alpha$  tri-peptide, Gly-Arg-Glu (GRE), and each had a different  $\omega$  tri-peptide. When the  $\omega$  tri-peptides were removed, there was no significant reduction in activity, suggesting that the ability of the GRE sequence to enhance fragment complementation did not depend on the presence of the  $\omega$  tri-peptide. Thus, the GRE  $\alpha$  tri-peptide produced a profound enhancement of the interaction-dependent activity, but it cannot substitute for the interaction. In fact, without the interaction the GRE tri-peptide does not seem to increase the background at all, thus it does not either accelerate refolding or stabilize the folded complex. The most likely effect of the GRE tri-peptide is to stabilize the  $\alpha 197$  fragment by interfering with loss of the fragment by amorphous aggregation. Since the  $\omega 198$  fragment is quite stable, but the  $\alpha 197$  fragment is somewhat less so, the latter is expected to be limiting for fragment complementation, and any stabilization of  $\alpha 197$  leading to an increase in its concentration would increase the steady state activity of the interaction-activated enzyme accordingly. Though the GRE tri-peptide could inhibit aggregation of  $\alpha 197$ , it apparently did not interfere with re-folding of the fragment complex. Since aggregate formation proceeds exponentially, it is exquisitely sensitive to small shifts in the inter-molecular association rate constants (Dobson, *Trends Biochem Sci* (1999) 24:329). Thus, even weak binding of the tethered tri-peptide to the interacting surfaces could effectively defeat inter-molecular aggregation. As the complementary fragments fold cooperatively into the active complex, however, the weakly bound tri-peptide would be readily stripped from its binding site by steric strain as the two become separated in the emerging native conformation. In this way the general ability of tethered small peptides to stabilize larger proteins without interfering with protein folding may be understood.

When the same random tri-peptide libraries were screened for fos/jun-mediated ampicillin resistance in the TG1 strain, five clones were recovered on 400  $\mu\text{g/ml}$  ampicillin. With the fos-jun interaction alone TG1 cells will not plate above 50  $\mu\text{g/ml}$  ampicillin.

Thus, as before, tri-peptides were selected which substantially increased the level of ampicillin resistance produced by the fos-jun interaction alone. This time four different  $\alpha$  tri-peptides were recovered, each with a different  $\omega$  tri-peptide.

5	Pairs	$\alpha$	$\omega$
	FHT400-1A1, -1B1	HSE (cat agt gag)	REQ (cgg gag cag)
	FHT400-2A1, -2B1	NGR (aat ggg cgg)	QGN (cag ggt aat)
	FHT400-4A1, -4B1	GRE (ggt cgg gag)	DGR (gat ggg agg)
10	FHT400-9A1, -9B1	EKR (gag aag cgt)	GRR (ggt agg agg)
	FHT400-10A2, -10B1	NGR (aat ggg cgg)	GNS (ggt aat agt)

GRE was selected again from the  $\alpha$  tri-peptide library. NGR was selected twice from the  $\alpha$  tri-peptide library, with two different  $\omega$  tri-peptides. In all cases, activation continued to be dependent on the fos-jun interaction. However, in contrast to the original GRE tri-peptide, activity was enhanced in all cases by the presence of the both the  $\alpha$  and  $\omega$  tri-peptides. Even the activity of the GRE tri-peptide was enhanced by the DGR tri-peptide on the  $\omega$  fragment. Also, the fragments were interchangeable to some extent. Different  $\alpha$  tri-peptides could be paired with different  $\omega$  tri-peptides. The fact that enhanced activity was still fully dependent on the heterologous interaction suggests that the primary effect of the peptides was protection of the fragments to which they were attached from aggregation, rather than stabilization of the final fragment complex. The latter would be expected to confer constitutive activity, independent of the heterologous interaction.

The GRE tri-peptide was also found to stabilize  $\alpha 197$  in *trans*. When the  $\alpha 197$ -fos and jun- $\omega 198$  fusions were co-expressed in the *E. coli* periplasm with the GRE tri-peptide fused to the N-terminus of thioredoxin via a Gly<sub>4</sub>Ser linker, the cells plated with 100% efficiency on 50  $\mu$ g/ml ampicillin, whereas cells expressing the  $\alpha 197$ -fos and jun- $\omega 198$  fusions either alone, without the GRE-*trxA* fusion, or with a different tri-peptide-*trxA* fusion, plated with only ~1% efficiency on 50  $\mu$ g/ml ampicillin. The GRE-*trxA* fusion conferred no resistance to ampicillin in the absence of the interacting helices, thus it does not stabilize the re-folded fragment complex, but rather it must stabilize the  $\alpha 197$  fragment since activity is limited by the amount of soluble  $\alpha 197$ . Since the GRE tri-peptide had the same stabilizing effect on  $\alpha 197$  fragment when a different carrier was used, its activity must be context independent. Thus, an 18 kDa enzyme fragment could be stabilized at least 100-fold by a tri-peptide selected from a random sequence library. As with the tethered

tri-peptide, the free GRE tri-peptide could inhibit aggregation of  $\alpha 197$  without apparently interfering with re-folding of the fragment complex. In this case, however, displacement of the tri-peptide would have been greatly assisted by the fact that the effective intramolecular concentrations of structural elements relative to one another would have been much higher than the tri-peptide concentration. In this way the general ability of small peptides to stabilize large proteins in *trans* without interfering with protein folding may be understood. This phenomenon is not widely appreciated, and in fact this may be the first demonstration that a functional protein could be deliberately stabilized by something as small as a tri-peptide.

10

### Example 7

#### Mutationally-Enhanced Fragment Complementation

The ability of tri-peptides to stabilize  $\beta$ -lactamase fragments and thereby to increase both the interaction-dependent activity and activation index of the TEM-1  $\alpha 197/\omega 198$  complex should be of great benefit for *in vitro* applications of  $\beta$ -lactamase fragment complementation, where utility is most limited by fragment instability. Thus, it was of interest to determine if a comparable stabilization of the  $\alpha 197$  fragment could be achieved by random mutagenesis and selection. To test this, the  $\alpha 197$  coding sequence was mutagenized by error-prone PCR (Cadwell and Joyce, 1995, *supra*). The PCR conditions of Cadwell and Joyce mis-incorporate nucleotides in an unbiased fashion at a rate of one mutation every  $\sim 150$  nucleotides. Since the  $\alpha 197$  coding sequence is actually about 520 nucleotides in length, and  $\sim 75\%$  of mutations change the encoded amino acids, less than three coding changes per molecule should be produced. About  $10^8$  clones of the  $\alpha 197$  mutant library were collected and co-expressed as the jun helix fusion with the fos helix fusion of wild-type  $\omega 198$ . The mutagenized  $\alpha 197$ -jun fusion was expressed from the pAE1 vector and the fos- $\omega 198$  fusion was expressed from the pAO1 phagemid vector (see Figure 6). When both constructs were co-expressed in strain DH5 $\alpha$  colonies were recovered in the presence of 600  $\mu\text{g/ml}$  ampicillin. Upon sequencing, two of three clones recovered (FI600-1 and -3) had the same sequence with two coding mutations, K55E (aag $\rightarrow$ gag) and M182T (atg $\rightarrow$ acg). The third clone (FI600-4) also had two coding mutations, one of which was shared with the other two (M182T), and the other of which, P62S (ccc $\rightarrow$ tcc), was proximal to the other mutation of the other clones.

20

25

30



Cells expressing either mutant consistently plated at  $>30\%$  efficiency on  $100\mu\text{g/ml}$  ampicillin, whereas cells expressing the wild-type  $\alpha 197$  plated at  $<10^{-6}$  colonies per cell on  $100\mu\text{g/ml}$  ampicillin, and  $\sim 30\%$  on  $25\mu\text{g/ml}$  ampicillin. However, for both mutants, plating efficiencies were just as high or higher in the absence of the heterologous interaction, i.e., with the jun helix removed. An exhaustive search for more mutations did not turn up any mutants with interaction-dependent activity. Thus, in contrast to the results obtained with random tri-peptides, where activation remained interaction-dependent, adaptive mutations of  $\alpha 197$  invariably eliminated interaction dependence. This may be understood as follows. The tri-peptides stabilized the fragments by reversibly interfering with aggregation. Reversibility allows them to inhibit aggregation without interfering with folding. However, mutations are not reversible in this sense. If aggregation is caused primarily by the inter-molecular formation of native folding contacts, disruption of these by mutation might be expected to interfere with folding. In fact, it may be thermodynamically impossible to stabilize the fragments by mutation without inhibiting the re-folding process required to form the active fragment complex. This is because the native folds of the fragments have too much exposed hydrophobic surface to be stable. Thus, mutations can only stabilize the fragments by stabilizing alternative folds, which minimize exposed hydrophobic surface. However, these alternative folds must be unfolded before the native folding pathway can proceed to the active complex, and the energy required for this process may be prohibitive.

Since most aggregation is driven by aggregation-prone intermediates in the folding pathway, the rate of aggregation is proportional to the lifetimes of such species. The effects of the break-point disulfide described above indicated that the fragments are capable of association and initiation of folding in the absence of the heterologous interaction, but that the folding process is aborted when the fragments are not held together in some way, such as by the heterologous interaction or by the formation of a disulfide at the break-point. In the absence of either of these the probability that the fragments will dissociate before folding is complete is proportional to the folding rate, which in turn is proportional to the lifetimes of the folding intermediates. Thus, if the most likely mechanism for mutational inhibition of aggregation is to destabilize folding intermediates, this would also accelerate folding and thereby reduce the probability that fragment dissociation would occur before folding were complete. In this way it may be understood why mutations which stabilize

the folded complex are more likely to be selected than mutations which stabilize the fragments, and why the former, but not the latter would give rise to constitutive, interaction-independent activity.

For the TEM-1  $\beta$ -lactamase of *E. coli*, the type member of the Class A penicillinases, fragments have been identified which can complement to form active enzyme when and only when the "break-point" termini of the fragments are fused to proteins or other molecules which interact with each other directly or preferably through a second molecule. Furthermore, the subject invention presents new methods whereby enzyme fragments capable of interaction-dependent complementation may be identified and modified specifically to confer dependence of their activity on the interaction of heterologous domains fused to the break-point termini. Ligand-activated or interaction-activated  $\beta$ -lactamases can be activated in multiple locations, including the bacterial periplasm, bacterial cytoplasm, eukaryotic cell cytoplasm, or *in vitro*. They are highly active against a wide variety of substrates, including antibiotics, chromogens, and fluorogens, as well as  $\beta$ -lactam pro-drugs, pro-antibiotics, and pro-nutrients, which can thus be used for both positive and negative viability selection and color selection. The utility of  $\beta$ -lactamase fragment complementation systems has been demonstrated for monitoring interactions between and among cell-surface receptors, antibodies, and random peptide libraries displayed on the surface of a natural protein.

### Example 8

#### Construction of a Human Peripheral Blood Lymphocyte Proteome Interaction Library.

The large number of functional interactions among both membrane-bound and secreted proteins of circulating immune cells include many which are yet to be discovered. For example, among the 150 or so CD antigens discovered so far, functions and ligands remain unknown for a substantial fraction (Ager *et al.*, in *Immunology Today Immune Receptor Supplement*, 2<sup>nd</sup> Ed. (1997). In addition, the highly combinatorial mechanisms by which signalling specificity is generated imply that many signalling proteins participate in multiple functional interactions, and that even the best known of these proteins may have ligands and functions which remain to be discovered. Thus, the functional interactions of the extra-cellular proteome of the circulating cells of the immune system represent a

potentially rich reservoir of pharmacological targets which are not readily accessible by currently available interaction mapping technologies. This proteome presents a unique opportunity to demonstrate the power of interaction-dependent  $\beta$ -lactamase fragment complementation systems for interaction mapping in that, while many important  
5 interactions remain to be discovered, many are already known by which the efficiency of the system can be gauged.

As discussed above, the activation index is the most important parameter of the interaction-dependent fragment complementation system for cleanly discriminating bona fide interactions from large pools of non-interacting protein pairs. Thus, for this  
10 application one would use the P174/N175 fragment pair of TEM-1  $\beta$ -lactamase ( $\alpha$ 174 and  $\omega$ 175) because with the break-point disulfide this pair has the largest activation index,  $\sim 10^7$ . It also has a robust specific activity, but this could probably be improved even further with some fragment-stabilizing tri-peptides, so one may first wish to insert the VRK  
15 or NNK tri-peptide library into the expression vectors between the break-point cysteines and the linkers (see Figure 6), and select for growth on 300-800  $\mu$ g/ml ampicillin. So long as the activation index is not compromised, higher specific activity conferred by fragment-stabilizing tri-peptides should allow weaker bona fide interactions in the expressed  
20 sequence libraries to confer selectable activity. In order to maximize the quality of the expressed sequence library, one might wish to subject the full-length cDNA library first to a normalization protocol to normalize the frequencies of rare and abundant sequences. From this normalized cDNA one would then prepare random primed cDNA by PCR, and size-select fragments  $> 200$  base-pairs to enrich the library for sequences which encode fragments which are at least the size of single protein domains. Finally the library could  
25 be run through a fold-selection protocol to enrich for coding sequences which are expressed in the correct reading frame and in register with autonomously-folding protein domains (AFD).

Rough microsomes, which are derived from membranes of rough ER and are therefore enriched in mRNA for secreted and membrane proteins, may be isolated from unfractionated lymphocytes from pooled human blood by sedimentation velocity in sucrose  
30 density gradients (Gaetani *et al.*, *Methods in Enzymology* (1983) 96:3; Natzle *et al.*, *J Biol Chem* (1986) 261:5575; Kopczynski *et al.*, *Proc Natl Acad Sci* (1998) 95:9973). Messenger RNA may then be purified from the rough microsomes using a commercially

available kit (e.g., Poly(A) Select, Promega, Inc., Madison, WI). A randomly-primed cDNA library is then made from the RNA template and cloned directionally. First-strand cDNA is made with AMV reverse transcriptase (RT) and random hexamer primers (Sambrook *et al.*, 1989, pp. 8.11-8.21). The primers contain a unique 5' extension with convenient restriction sites for ligation into the  $\beta$ -lactamase  $\alpha$  and  $\omega$  fusion expression vectors. The template is destroyed by the RNaseH activity of AMV RT and the unused primers are removed using a spun column. The second strand is then made with the Klenow fragment of DNA polymerase I and random hexamer primers containing a different unique 5' extension with a different restriction site for insertion into the expression vectors.

After removal of unused primers, the cDNA is PCR-amplified with primers corresponding to only the unique sequence on each original primer (Dieffenbach and Dveksler, in *PCR Primer: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1995), so that the majority of amplified fragments have the correct orientation for expression in *E. coli*. The product is then normalized by exhaustive hybridization to a limiting amount of human genomic DNA immobilized on magnetic beads (Kopczynski *et al.*, 1998, *supra*). Since coding sequences are naturally normalized in genomic DNA, cDNA recovered from the genomic DNA hybrids should be normalized. After a final amplification, the PCR product is size selected by centrifugal gel filtration on Sephacryl S-400 spun columns for fragments  $> \sim 200$  bp. The cDNA is then digested with appropriate restriction enzymes and ligated into the interaction-dependent  $\beta$ -lactamase  $\alpha 174$  and  $\omega 175$  fusion expression vectors, which are essentially the same as those shown in Figure 6, except for some modifications required for fold selection. The vectors and protocol for fold selection and interaction mapping of the cDNA library are illustrated in Figure 9.

For convenient fold selection, both vectors for expression of the library as  $\alpha$  and  $\omega$  fusions are compatible phagemids. In addition, a peptide epitope tag, such as the well-known 12-mer derived from the *c-myc* oncogene (Hoogenboom *et al.*, 1998, *supra*) is encoded at the C-terminus of the cDNA, or expressed sequence (ES) library in the  $\alpha$ -fusion vector, and at the N-terminus of the ES library in the  $\omega$ -fusion vector. When co-expressed with an anti-tag scFv, such as the anti-myc 9E10 scFv (Hoogenboom *et al.*, 1998, *supra*) fused to the other  $\beta$ -lactamase fragment, each fusion library can be enriched for clones which express autonomously folding domains (AFD) in the correct reading frame. The principle of the selection is that only fragments which can fold into their native

conformations will be stable enough to support selectable levels of  $\beta$ -lactamase fragment complementation driven by the tag-anti-tag interaction.

The normalized cDNA library-vector ligation products are transduced into *E. coli* strain TG-1 by high-voltage electroporation (Dower *et al.*, *Nucleic Acids Res* (1988) 16:6127), and plated onto the minimum ampicillin concentration on which non-interactors are known to plate with efficiencies of  $\leq 10^{-3}$  since at least a 100-fold excess of non-AFD-encoding fragments is expected in the libraries. For the  $\alpha 174/\omega 175$  system, the recommended ampicillin concentration would be  $\sim 25$   $\mu\text{g/ml}$ . Since there is not likely to be more than  $10^4$  secreted or membrane protein genes expressed in PBLs, and the frequencies of expressible AFDs may be in the range of  $10^{-2}$  per gene, one should collect at least  $10^7$  clones of each library to insure representation of all expressible extra-cellular AFDs.

Once the normalized ES libraries have been enriched for AFD-encoding clones, the libraries can be rescued as filamentous phage by high-multiplicity super-infection of at least  $10^8$  cells of each library with the helper phage M13K07 (Sambrook *et al.*, 1989, pp. 4.17-4.19). After overnight growth in suspension the library phage are recovered from the culture supernatant by precipitation with polyethylene glycol, and reconstituted in phosphate-buffered saline. The library phage stocks may be stored frozen in 15% glycerol. Fresh *E. coli* TG-1 cells may then be co-infected with a high-multiplicity of each phage library and plated onto a concentration of ampicillin on which the activation index of the system is known to be maximal. For the  $\alpha 174/\omega 175$  system, 100  $\mu\text{g/ml}$  ampicillin is optimal, since the activation index is at least  $10^7$  and the fos-jun interaction-mediated plating efficiency is at least 50%. At least  $10^{14}$  transforming units of each fusion library phage should be used to infect at least  $10^{12}$  log phase TG-1 cells to insure that most of the possible pair-wise combinations of  $10^6$  clones of each AFD library are present in the doubly infected cell population before selection. After a one-hour adsorption at  $10^9$  cells per ml, the cells are washed, resuspended in fresh medium, and incubated for another hour with gentle shaking to allow the phagemid genes to express. The cells are then concentrated and plated on 100 large petri dishes (150 mm dia.) containing solid LB medium containing 1 mM IPTG and 100  $\mu\text{g/ml}$  ampicillin. A small aliquot is plated on chloramphenicol and kanamycin to determine the number of co-transformants.

Since  $\sim 10^{10}$  cells are being seeded onto each plate, it is possible that the interaction

frequency might be high enough for the plates to overgrow. This would take at least  $10^4$  clones per plate. In this case, all of the selected clones would have to be recovered by scraping and replated at lower densities. If a large number of clones is recovered, at least 100 should be replated anyway to determine the background frequency due to ampicillin escapes. From those that breed true, each candidate interactor should be recovered and tested for interaction with an unselected partner. Selected pairs should be sequenced and BLAST-searched for homology to known genes (Altschul *et al.*, *J Mol Biol* (1990) 215:403; Altschul *et al.*, *Nucleic Acids Res* (1997) 25:3389). A large number of interactions among secreted and membrane proteins of immune cells are already known, such as the B-cell co-activation antigen, CD40 and its T-cell ligand, CD40L, and the T-cell activation antigens B7.1 and B7.2 and their ligands CD28 and CTLA4. Labeled oligonucleotide hybridization probes may be prepared for these known interactions, and colony lifts of the entire interaction library may be probed to see what fraction of expected interactors are actually represented in the library. Interaction partner sequences from positive clones may be recovered, and homology searched to determine if known or new interactors have been identified. Colonies expressing bona fide interactions may be grown up and stored indefinitely in 15% glycerol at  $-70^{\circ}\text{C}$ , pending further characterization or use for e.g., drug screening.

### Example 9

#### Construction of an Intra-Cellular Signal Transduction Biosensor.

Interaction-dependent  $\beta$ -lactamase fragment complementation systems can be adapted for activation or inactivation by virtually any post-translational modification that occurs naturally in cells. As a result they may be deployed intra-cellularly as biosensors to monitor the activity of any process which is regulated by post-translational modification. A major class of such processes is phosphorylation-regulated signal transduction pathways. Phosphorylation-regulated intermediates are obligatory components of most processes by which cells respond to extra-cellular conditions or messenger molecules by altering gene expression. Cellular responses to extra-cellular signals may be fall into three general categories, growth, survival, and differentiation. A ubiquitous component of neoplastic transformation is the deregulation of growth control signaling, often accompanied by the deregulation of survival signalling as well. This often occurs by over-expression of

phosphorylation-regulated signal transducers, or by mutational disabling of phosphorylation-mediated regulation. Thus, most so-called oncogenes are phosphorylation-regulated growth signal transducers, which become over-expressed or mutated to constitutive activity in cancer cells.

5 The Her-2/neu oncogene is a 185 kDa Type I transmembrane receptor tyrosine kinase, which is a member of the epidermal growth factor receptor (EGFR) family. This growth factor receptor is over-expressed in particularly aggressive adenocarcinomas of epithelial origin in a number of tissues, notably breast. When normally expressed, Her-2/neu hetero-dimerizes with other EGF-family receptors when they are ligated by growth  
10 factor. This leads to cross phosphorylation of multiple tyrosines on the cytoplasmic domains of the receptors. Phosphorylation of tyrosine 1068 (Tyr1068) on Her-2/neu leads via phospho-tyrosine-binding accessory proteins and guanosine nucleotide exchange factors to activation of p21<sup>ras</sup>, and thence to activation of cell division via the MAP kinase cascade. When Her-2/neu is sufficiently over-expressed, the background level of ligand-  
15 independent EGFR hetero-dimerization rises to a level which is in turn sufficient to maintain constitutive mitogenic signaling even in the absence of growth factor, leading to the characteristically uncontrolled growth of tumor cells. Thus, there is much interest in finding drugs which can block the activation of Her-2/neu, particularly in a manner which can prevent constitutive signaling in tumor cells without blocking EGF signalling in normal  
20 cells.

A cell-based biosensor, which produces a readily detectable and quantifiable signal when Her-2/neu activation is blocked, would be particularly useful for high-throughput screening of chemical libraries for compounds with anti-breast tumor potential. Such a biosensor may be set up with a  $\beta$ -lactamase fragment complementation system as follows.  
25 The  $\omega$  fragment could be fused via flexible linker to the C-terminus of Her-2/neu, which is proximal to the Tyr1068 substrate of the receptor kinase. The  $\alpha$  fragment could then be fused to a binding protein, such as a scFv or VL, which binds to the Tyr1068 region of the receptor only when Tyr1068 is unphosphorylated. Since Tyr1068 is mostly phosphorylated in Her-2/neu over-expressing cells, especially in the presence of EGF,  $\beta$ -lactamase  
30 activation would be minimal. However, in the presence of an inhibitor of Her-2/neu activation, the proportion of unphosphorylated Tyr1068 would rise, recruiting the  $\alpha$ -Tyr1068 binder fusion to the receptor where  $\alpha$ - $\omega$  complementation would increase  $\beta$ -

lactamase activity in the cells. In the presence of a fluorogenic  $\beta$ -lactamase substrate, inhibitors of Her-2/neu activation could be readily identified by increasing fluorescence in a matter of minutes, since dephosphorylation of Tyr1068 occurs rapidly upon inhibition of the Her-2/neu kinase activity.

5 For intra-cellular biosensors both maximum activity and the activation index will be important. However, for all five of the best TEM-1 fragment pairs the activation index is expected to depend almost entirely on the difference in the affinity of the binder for Tyr vs phospho-Tyr. Thus, the fragment pair with the highest activity, i.e., G253/K254 ( $\alpha$ 253 and  $\omega$ 254), would be preferred, especially since for intra-cellular applications the break-  
10 point disulfide cannot be used. It may be possible to increase the intra-cellular activity of  $\alpha$ 253/ $\omega$ 254, if desired, by selecting one or two fragment stabilizing tri-peptides, as described above.

15 The first step in developing the Her-2/neu inactivation biosensor would be to obtain a Tyr1068-binding protein. This could be accomplished by inserting the coding sequence for the substrate peptide, PVPEYINQS, into the active site of thioredoxin, between G33 and P34, flanked by short flexible linkers such as PGSGG to minimize structural constraints on the peptide, which does not require a rigid structure for binding to its natural ligand, the Grb2 SH2 domain. This Tyr1068 trxpep can then be fused via a (Gly<sub>4</sub>Ser)<sub>3</sub> linker to the N-terminus of  $\omega$ 254, and co-expressed in *E. coli* TG-1 cells with a scFv  
20 library of at least  $10^8$  clones, or a VL library of at least  $10^6$  clones fused to the C-terminus of  $\alpha$ 253 via the (Gly<sub>4</sub>Ser)<sub>3</sub> linker. Since the Tyr1068-binder is being selected for deployment in the mammalian cell cytoplasm, it might be prudent to perform the selections in the *E. coli* cytoplasm. For this purpose the vectors in Figure 6 could be used with the signal peptides removed. Then a chromogenic substrate such as nitrocefin ( $\lambda_{\max}$  = 485  
25 nm;  $\epsilon$  = 17,420 M<sup>-1</sup> cm<sup>-1</sup>; McManus-Munoz and Crowder, *Biochemistry* (1999) 38:1547) would be used to select Tyr1068-binders by color. By plating at least  $10^6$ - $10^8$  transformants at moderate to high stringency, i.e., on decreasing concentrations of the substrate, it should be possible to identify binders with sub-micromolar affinities since Tyr is the most common amino acid in high-affinity protein-protein interfaces. Such affinities  
30 will be desirable for maximum discrimination between Tyr and phospho-Tyr. Selected Tyr1068-binders must be tested for inhibition by phosphorylation of the Tyr. This can



2/12/01  
2/12/01  
2/12/01  
easily be accomplished by expressing the vectors in isogenic cells which over-express a broad spectrum Tyrosine kinase (TKX1 cells, Stratagene, Inc., La Jolla, CA).

Once a suitable phosphate-sensitive Tyr1068-binder has been identified, the entire coding sequence for the  $\alpha$ 253 - Tyr1068-binder fusion may be subcloned into a mammalian expression vector, such as the pCMV-Tag vectors (TKX1 cells, Stratagene, Inc., La Jolla, CA) for expression in mammalian cells from the cytomegalovirus promoter.

The  $\omega$ 254 fragment must be expressed as a fusion to the C-terminus of the Her-2/neu cytoplasmic domain, which contains Tyr1068. The coding sequence of the 1210-residue EGF receptor (Genbank accession no. X00588; Ullrich *et al.*, *Nature* (1984) 309:418) may be used as it is operationally identical to Her-2/neu, and its Tyr1068 will become

phosphorylated under the same conditions of over-expression and/or growth factor ligation in tumor cells. When fused to the C-terminus of EGFR via the (Gly<sub>4</sub>Ser)<sub>3</sub> linker, the 35-residue  $\omega$ 254  $\beta$ -lactamase fragments will be only 152 residues away from Tyr1068. Both

the EGFR- $\omega$ 254 fusion and the  $\alpha$ 253-Tyr1068-binder fusion may be expressed from the same vector from a dicistronic mRNA. This is accomplished by inserting an internal

ribosome entry site (IRES; Martinez-Salas, *Curr Opin Biotechnol* (1999) 10:458) between the termination codon of the upstream cistron and the initiation codon of the downstream cistron. This will allow both proteins to be made simultaneously from the same mRNA.

The vector may be introduced into the tumor cell line by cationic liposome-mediated transfection, using e.g., lipofectamine (Gibco-BRL, Gaithersburg, MD) according to the protocol in the product literature. Operation of the biosensor may be tested in transiently transfected cells, and if operational, stable transformants may then be isolated by selection for long term antibiotic resistance. Multiple free-diffusible chromogenic and fluorogenic substrates are available for continuous monitoring of  $\beta$ -lactamase activity. Operationally,

the  $\omega$ 254 fragment will be anchored to the plasma membrane at the C-terminus of the cytoplasmic domain of the receptor near Tyr1068, and the  $\alpha$ 253 fragment will be free in the cytoplasm as the Tyr1068-binder fusion. ATP-analog tyrosine kinase inhibitors are available commercially and can be used as positive controls for inhibitor selection, and to determine the signal increment from fully-activated to fully-inhibited EGFR.

### Example 10

#### A Fragment Complementation System for Neomycin Phosphotransferase.

Enzyme fragment complementation systems may also be useful for selection for the simultaneous incorporation of multiple genetic elements into the same cell or organism.

5 For example, the production of secretory IgA antibodies in plants requires the introduction of four different genes into the same plant. For practical reasons this requires the introduction of at least two and preferably three different DNA molecules. For the production of genetically stable transgenic plants, each DNA molecule must carry its own selectable marker. The use of multiple antibiotic selection systems on the same  
10 transformants is cumbersome and inefficient, as the overall false positive and false negative rates tend to scale as the product of the rates for the individual antibiotics. Thus, two- or three-piece fragment complementation systems for a single antibiotic offer a distinct

advantage over multiple antibiotic selection.

For a two fragment system, dependence of activation on the interaction of  
15 heterologous domains is not necessary. However, for simultaneous selection of triple transgenics, complementation of the enzyme fragment pair must be dependent on a heterologous interaction mediated by a free ligand, analogous to the activation of  $\beta$ -lactamase by the tri-molecular interaction of  $\alpha$ 197-jun, scFv- $\omega$ 198, and CD40-fos, as described above. For these applications, the most important parameter is the maximum  
20 activity of the reconstituted enzyme, which is a function of both the specific activity and the efficiency of complementation. The activation index is not relevant because each fragment alone will have essentially no detectable activity, providing a background of zero.

Thus, to insure recovery of the most competent fragment pairs for intra-cellular activity, the fos and jun interactors should be used with tri-peptide libraries between the break-  
25 points and the (Gly<sub>4</sub>Ser)<sub>3</sub> linkers. The tri-peptide libraries will provide stabilizers for each fragment so that the selection will be biased toward the fragments producing the highest specific activities. For two-trait selection applications, i.e., bi-molecular selections, where a heterologous interaction is not required, specific activity may be increased further by mutagenesis and selection for fold accelerating mutations. For three-trait selection  
30 applications, selected fragment pairs will have to be tested for dependence on the heterologous interaction. In this case, the activation index will be of some importance, but

as with in vitro applications a modest index of 1000 will be more than adequate for clean selections.

Neomycin phosphotransferase II (NPTII; Genbank accession no. M77786) is a 267-amino acid enzyme from *E. coli* which inactivates aminoglycoside antibiotics such as neomycin and kanamycin by phosphorylation from ATP. NPTII is widely used as a selectable marker for plant and animal cell transformation. Thus, fragment complementation systems for NPTII would be particularly useful for facile generation of multiple-trait plant and animal transgenics. The three-dimensional structure of NPTII is not known, and its homology to known structures is too low for reliable prediction.

However, as described above, empirically-derived neural net algorithms are available which allow fairly accurate prediction of secondary structure and solvent exposure for any protein sequence. The best of these algorithms is the PredictProtein program of Rost and Sander (1993, 1994, *supra*). Application of this program to the protein sequence of NPTII reproduced the result shown in Figure 10. Ten regions of the sequence have been predicted to have little secondary structure and to be exposed to solvent, and therefore to be potential sites for productive fragmentation. Fragment pairs corresponding to breakage in the center of each of these ten regions, or at two equally-spaced sites in the longer regions, may be generated by PCR with appropriate primers, and subcloned into vectors like those illustrated in Figure 6 for expression as the fos and jun helix fusions with intervening linkers. The vectors would differ from those in Figure 6 in not encoding signal peptides, and the pAO1 vector would have ampicillin resistance instead of kanamycin resistance. Also, the vectors should contain VRK or NNK random tri-peptide-encoding sequences between the cloning sites for the enzyme fragments and the (Gly<sub>4</sub>Ser)<sub>3</sub> linkers.

The PCR product for each fragment is restriction digested and ligated into the appropriate vector,  $\alpha$  fragments into the pAE1-type vector and  $\omega$  fragments into the pAO1-type vector. The ligation products are then introduced into TG-1 cells by high-voltage electroporation, and plated onto chloramphenicol or ampicillin. At least  $10^4$  transformants should be collected for each fragment. Also, kanamycin sensitivity should be determined for each fragment library, both to prevent false positives and to determine the minimum quantitatively selective kanamycin concentration. This should be the concentration on which single fragment plating efficiencies are  $< 10^{-6}$ , since the frequencies of the fragment-stabilizing peptides could be this low. Since  $\sim 10^8$  co-transformants will be needed for

each fragment pair for complete coverage of the tri-peptide libraries, quantitative phage infection should be used to combine the two libraries for each fragment pair. This is accomplished by rescuing the  $\omega$ -fragment libraries (in the pAO1-type phagemid vector) as phage using M13K07 helper phage as described above. For facile quantitative infection at least  $10^9$  cells bearing each  $\alpha$  fragment library should be inoculated with at least  $10^{11}$  phage bearing the corresponding  $\omega$  fragment library. After one-two hours in suspension culture with gentle shaking to allow phage adsorption, penetration, and initiation of gene expression, the cells of each fragment pair are centrifuged, washed, and plated onto ten 150-mm dishes containing solid LB medium with the minimum quantitatively selective concentration of kanamycin.

After overnight growth at 37°C, all kanamycin-resistant colonies may be pooled and re-plated onto increasing concentrations of kanamycin to identify those tri-peptide/fragment pair combinations producing the highest levels of kanamycin resistance. As many of the most active clones as necessary should be tested for dependence of activity on the fos-jun interaction. This can most easily be accomplished by removing one of the helices by restriction digestion at sites in the gene construct included for this purpose. The digestion products are then re-ligated, re-transformed into TG-1 cells, and replated on kanamycin. As explained above activation indexes of 1000 are more than adequate, so the most active pairs with indexes of at least 1000 would be optimal. For tri-molecular activation in the cytoplasm, two hetero-dimerizing helix pairs may conveniently be used, such as the parallel-binding helices from *fos* and *jun* as described above, and the anti-parallel-binding helices from yeast DNA topoisomerase II (TopII; Berger *et al.*, *Nature* (1996) 379:225). One of each helix pair would be fused to an NPTII fragment, and the other two helices would be fused to each other, so that the NPTII fragments would only come together when the 2-helix fusion was present to form the tri-molecular complex. For example, an  $\alpha$ -TopIIN fusion and a fos- $\omega$  fusion could only be brought together and activated by a jun-TopIIC fusion. Genes encoding each of the three fusions could then be distributed among three different DNA constructs which also encode genes of interest. In this way eukaryotic cells could be transformed with a mixture of the three different constructs and selected for the simultaneous presence of all three genes in the same cell simply by selection for growth on a single antibiotic.

### Example 11

#### Target-Activated Enzyme Prodrug Therapy.

Antibody-directed enzyme prodrug therapy (ADEPT) is a promising anti-cancer chemotherapeutic strategy which takes advantage of the catalytic power of enzymes to amplify the cytotoxicity-targeting power of tumor-specific antibodies. Enzymes are concentrated at the tumor site when administered as conjugates of tumor-specific antibodies. After unbound conjugate has cleared from the circulation, prodrugs may be administered which are relatively non-toxic until activated by the tumor-bound enzyme, whereupon the cytotoxic product may accumulate at the tumor site to concentrations which would be unattainable by parenteral administration of the drug without excessive toxicity. Enzymes such as  $\beta$ -lactamase have been chemically or genetically conjugated to tumor-targeting antibodies and used with  $\beta$ -lactam derivatives of anti-tumor drugs such as cephalosporin mustards and anthracyclines to achieve promising anti-tumor effects in animals. The efficacy of ADEPT is limited, however, by the need for unbound conjugate to clear the circulation before the prodrug can be administered. By the time the circulating conjugate is depleted to the threshold below which systemic activation of the prodrug would produce acceptable levels of toxicity, so much of the conjugate has been lost from the tumor that efficacy is often seriously compromised.

This problem may be overcome by using an interaction-dependent  $\beta$ -lactamase fragment complementation system with tumor targeting antibodies. When fused to single-chain antibody fragments (scFv) which recognize non-overlapping epitopes on tumor markers, the  $\beta$ -lactamase fragments can localize to the tumor and reconstitute sufficient  $\beta$ -lactamase activity on the tumor cell surface to produce high levels of tumor-localized cytotoxicity from  $\beta$ -lactam prodrugs. The great advantage of such a system is that prodrug activation cannot occur in the general circulation or anywhere the tumor marker is not encountered, so that the prodrug may be administered either simultaneously with high doses of the scFv-fragment fusions, or at the point of highest tumor load of the fragments, without regard for the circulating levels of the fragments which would be completely inactive.

As an example, the construction and purification of fusions of interaction-dependent  $\beta$ -lactamase fragments with scFv which bind non-overlapping epitopes on the human breast tumor marker Her-2/neu is described. One may then determine the kinetics of

reconstitution of  $\beta$ -lactamase activity on the surface of Her-2/neu – expressing SKOV3 human ovarian cancer cells. Under conditions of optimum loading, killing of the cells may then be assessed for different cephalosporin prodrugs as a function of concentrations known to be limiting in vivo. The resulting Tumor-Activated Enzyme Prodrug Therapy (TAcEPT) system may then be tested for its ability to ablate SKOV3 and other Her-2/neu-expressing human tumors in severe combined immuno-deficient (scid) mice. Once the efficacy and safety of the system has been demonstrated in animal models, toxicity and efficacy trials may be initiated in human breast cancer subjects.

The requirements for therapeutic use of  $\beta$ -lactamase fragment complementation systems are similar to those for in vitro use in general. The most important parameters are specific activity and fragment stability, while activation indexes above 1000 confer little additional efficacy. Thus, the  $\alpha$ 253/ $\omega$ 254 would be the recommended fragment pair for this application because it has the highest interaction-dependent specific activity, the fragments are moderately stable, and its activation index is more than adequate. However, the stability of the  $\alpha$ 253 fragment could probably be improved by a custom fragment-stabilizing tri-peptide. Thus, before setting up the tumor-activated system, one might first subclone a degenerate sequence encoding the VRK or NNK tri-peptide library into the  $\alpha$ 253 expression construct between the break-point cysteine and the linker (see pAE1 in Figure 6).  $\alpha$ 253-stabilizing tri-peptides would then be selected by plating at least  $10^4$  library transformants on increasing ampicillin from 400 to 1000  $\mu$ g/ml, since  $\alpha$ 253/ $\omega$ 254 plates quantitatively on 400  $\mu$ g/ml even without a stabilizing peptide, and wild-type TEM-1  $\beta$ -lactamase does not plate on more than 1000  $\mu$ g/ml when expressed under these conditions.

11a. Expression of TEM-1  $\beta$ -lactamase H25-G253 ( $\alpha$ 253) and K254-W288 ( $\omega$ 254) fragments as fusions to scFv against non-overlapping epitopes on the Her-2/neu human breast tumor marker.

The tumor activation mechanism for these fragments may employ two scFvs such as those described by Schier *et al.* (*Gene* (1996) 169:147), which were derived from a phage display library of a human non-immune repertoire (Marks *et al.*, 1991) by panning against a recombinant fragment comprising the extra-cellular domain (ED) of Her-2/neu. These two scFv, appear to recognize non-overlapping epitopes, since they do not compete for

binding to the Her-2/neuED by ELISA. The affinity of one of these scFv was improved to sub-nM Kd in vitro (Schier *et al.*, 1996, *supra*), and similar improvements in the other could be made using the same methods (Balint and Larrick, *Gene* (1993) 137:109). The coding sequences for the scFv may be subcloned into the  $\beta$ -lactamase  $\alpha$  and  $\omega$  fusion production vectors, p $\beta$ lac $\alpha$  and p $\beta$ lac $\omega$ , shown in Figure 11. These vectors are derived from pET26b (Novagen), and have convenient restriction sites for insertion of both scFv and  $\beta$ -lactamase fragment sequences. Each fusion protein is inducibly expressed (IPTG) from the strong phage T7 promoter under the control of the *lac* repressor. Each primary translation product contains a pelB signal peptide for secretion into the bacterial periplasm and a C-terminal His<sub>6</sub> tag for one-step purification from osmotic shock extracts by immobilized metal ion affinity chromatography (IMAC, Janknecht *et al.*, *Proc Natl Acad Sci* (1991) 88:8972). The yield of each fusion protein can be optimized primarily by manipulation of the inducer concentration and the growth temperature.

Each scFv may be expressed as both  $\alpha$ - and  $\omega$ -fusions to determine which arrangement(s): (1) support the highest binding activity; (2) support the highest enzymatic activity; and (3) support the highest yields. Initially, expression may be optimized by the criterion of silver-stained PAGE. Then fusion proteins should be purified from osmotic shock extracts (Neu and Heppel, 1965, *supra*) by IMAC. The purified fusion proteins may be tested for binding to an immobilized recombinant fusion of the Her-2/neu extra-cellular domain (ED) to a stabilizing immunoglobulin domain (Ig) by ELISA using an anti-His<sub>6</sub> tag antibody (Qiagen). The purified fusion proteins may then be tested for reconstitution of  $\beta$ -lactamase activity on immobilized rc- Her-2/neu ED-Ig using a chromogenic substrate, nitrocefin ( $\lambda_{max}$  = 485 nm;  $\epsilon$  = 17,420 M<sup>-1</sup> cm<sup>-1</sup>; McManus-Munoz and Crowder, 1999, *supra*). Immobilized BSA may be used as the negative control.

11b. Determination of the kinetics of specific  $\beta$ -lactamase activation by binding of  $\beta$ -lac $\alpha$ / $\omega$ -scFv fusions to immobilized recombinant antigen.

One may determine  $\beta$ -lactamase activity quantitatively as a function of binding of the fusion proteins to the immobilized antigen. This rate may then be compared to that obtainable with intact  $\beta$ -lactamase fused to the same scFv as an indication of how much activity may be localized on a tumor compared to an established vehicle, i.e., an antibody- $\beta$ -lactamase conjugate.

First, conditions are established for saturating the antigen with one of the scFv- $\beta$ -lac fragment fusion proteins. The wells of microtiter plates are coated with antigen, and exposed to increasing amounts of the first scFv-fragment fusion until the ELISA signal plateaus. At this level, i.e., saturating amounts of the first fusion protein, increasing amounts of the second fusion is added. After binding and washing,  $\beta$ -lactamase activity is determined spectrophotometrically after a 30' incubation with excess nitrocefin. If the assay is performed in triplicate,  $V_{max}$  should be a more or less linear function of the concentration of the second fusion. As the amount of second fusion is increased, at some point  $V_{max}$  should plateau. The amount of the second fusion bound can be determined by ELISA, and a relative specific activity ( $k_{cat}^{rel}$ ) may be computed for the fragment-reconstituted  $\beta$ -lactamase. The  $K_M$  may be estimated in solution with saturating antigen and saturating first fusion and limiting amounts of the second fusion. A range of nitrocefin concentrations is added and the initial rates of change of absorbance at 485 nm is measured as a function of second fusion concentration. The  $K_M$  is then computed from standard regression analysis.

To compare with intact  $\beta$ -lactamase, a fusion of intact  $\beta$ -lactamase to the second scFv may be prepared. This is then added in increasing amounts to antigen-coated wells which had been saturated with the first fusion as had been done before. Again,  $V_{max}$  should be a more or less linear function of the amount of intact  $\beta$ -lactamase fusion and should plateau at saturation. At each point, the amount of intact  $\beta$ -lactamase fusion bound, as determined by ELISA, should be comparable to the amount of the second fragment fusion bound, and the ratio of  $V_{max}$  should reflect the ratio of specific activities of the intact and fragment-reconstituted  $\beta$ -lactamases. For comparison, the  $K_M$  should be estimated as described above for the fragment-reconstituted enzyme. The TEM-1  $\alpha 253/\omega 254$  fragment complex is expected to have a maximum activity ( $k_{cat}$ ) near that of the intact enzyme. If the  $K_M$  are also comparable, activities on a tumor up to 100-fold higher at the peak of prodrug activation than with the conventional antibody- $\beta$ -lactamase fusion might be expected, which may have 1% or less of its peak activity left when the unbound fusion has cleared the circulation enough to allow prodrug administration.

11c. Determination of killing kinetics of Her-2/neu-expressing SKOV3 ovarian carcinoma cells by scFv-mediated  $\beta$ -lac $\alpha/\omega$  activation of cephalosporin prodrugs.



The arrangement(s) of scFv- $\beta$ -lactamase fragment coupling which produce(s) the highest specific  $\beta$ -lactamase activities on immobilized antigen may then be tested for activation of  $\beta$ -lactamase activity in the presence of human tumor cells expressing the Her-2/neu antigen. Cell killing may be assayed using any of the three cephalosporin prodrugs shown in Figure 5. The fragment-reconstituted activity may again be compared with the intact  $\beta$ -lactamase activity, this time with respect to tumor cell killing. Such results should indicate the dose range which may be required to show a significant anti-tumor effect in animals, which will be the next step in preclinical evaluation of the tumor-targeted  $\beta$ -lactamase.

The SK-OV-3 line of human ovarian adenocarcinoma cells (ATCC) may be seeded in 6-well tissue culture plates at  $3 \times 10^5$  cells per well in Dulbecco's Minimum Essential Medium (DMEM) supplemented with 10% fetal calf serum (FCS), and allowed to grow to confluency at  $37^\circ\text{C}$  in 10%  $\text{CO}_2$ . The saturability of both Her-2/neu epitopes on the cells may be determined with increasing amounts of intact  $\beta$ -lactamase fused to each scFv, as determined spectrophotometrically after nitrocefin hydrolysis. The  $V_{\max}$  of the fragment-reconstituted enzyme may then be determined on the cells with saturating concentrations of both fusions and nitrocefin. It would be expected to conform to the predicted activity based on the maximum intact  $\beta$ -lactamase activity and the ratio of  $V_{\max}$  observed on the immobilized recombinant antigen. The sensitivity of the cells to any of the three prodrugs shown in Figure 5 may be determined essentially as described by Marais *et al.* (*Cancer Research* (1996) 56:4735) with and without the intact  $\beta$ -lactamase-scFv fusions and the  $\alpha/\omega$  fragment-scFv fusions under saturating conditions. The prodrugs are dissolved in DMSO and diluted into DMEM/FCS to a range of concentrations immediately prior to use. One ml is added to each well and the cells are incubated overnight. The cells are then washed, trypsinized, and viability is determined by dye exclusion. Aliquots are then seeded into fresh dishes. After four days of growth, cell viability is assessed by incorporation of [ $^3\text{H}$ ] thymidine as determined by liquid scintillation counting of acid insoluble material. The results are expressed as percentage of untreated control cells. Again, the relative cytotoxicities of the prodrugs with the  $\beta$ -lactamase fragment system may be compared to those of the intact  $\beta$ -lactamase fusions, particularly at the lower prodrug concentrations where second order rate constants ( $k_{\text{cat}}/K_M$ ) may be important, to give an indication of the potential increase in efficacy of TAcEPT over conventional ADEPT in vivo.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporate by reference.

The invention now having been fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	